ELECTROMAGNETIC THEORY AND SIMULATIONS OF OPTICAL INTERACTIONS WITH METAL NANOSTRUCTURES

by

Yonggang Xi

B.S. in Physics, Fudan University, 2001

M.S. in Physics, Fudan University, 2004

Submitted to the Graduate Faculty of

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Yonggang Xi

It was defended on

Nov. 5th, 2010

and approved by

Joel Falk, Professor, Department of Electrical and Computer Engineering

Mahmoud El Nokali, Associate Professor, Department of Electrical and Computer Engineering

Guangyong Li, Assistant Professor, Department of Electrical and Computer Engineering

Rob Coalson, Professor, Department of Chemistry

Dissertation Director: Hong Koo Kim, Professor, Department of Electrical and Computer Engineering

ELECTROMAGNETIC THEORY AND SIMULATIONS OF OPTICAL INTERACTIONS WITH METAL NANOSTRUCTURES

Yonggang Xi, PhD

University of Pittsburgh, 2010

Interaction of light with a nanoapertured metal film has been the subject of extensive study because it produces many interesting phenomena, such as “enhanced” transmission of light through a nanohole array or shaping the spatial or spectral profiles of the transmitted light. The richness of the phenomena stems from the complexity of the way that light interacts with the nanostructures formed in the metal film. Surface plasmons (SPs), collective oscillation of electrons carrying the electromagnetic energy in the form of photons trapped at a metal/dielectric interface, can effectively mediate the interactions between metal nanostructures. Unlike the dielectric case, a metal wedge structure can also efficiently interact with free-space radiation, diffracting an incident light and/or coupling the light into surface plasmons (vice versa, decoupling surface plasmons into free-space radiation). We show that the diffraction by a metal corner plays an essential role in exciting surface plasmons and shaping energy flow distributions (enhancement or depletion). The phase relationship of the boundary diffraction and planar incident waves is extracted from measurement and simulation results.

A single nanoaperture formed in a metal film is a simple and yet one of the most fundamental structures that can be viewed as a basic building block of aperture-based nano-plasmonic structures. In this study we have investigated the characteristic evolution of optical wavefronts emanating from a nanoslit formed in a thin silver film. A planar wave, directly transmitted through the thin metal film, was used as a reference in forming an interference

pattern with the slit-transmitted free-space radiation and surface plasmons, and a scanning probe technique was employed in imaging the interference pattern in the near- to far-field regimes. Both the amplitude and phase information of the slit-transmitted waves with respect to the direct film-transmitted wave were extracted from the experimental data, and the results are compared with the analytical and numerical simulation results. The near- to far-field imaging of optical wavefronts is expected to be important in designing advanced nano-optic and plasmonic structures where precise control of both phase and amplitude of an optical signal is essential. We have also investigated grating diffraction with order-selection capability. This method offers a promising approach to accessing angular ranges that have not been reachable in conventional optics and to overcoming the limits of conventional refractive optics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This thesis would not have been possible without the generous support, helpful guidance and good wishes from many people. First and foremost, I would like to express my eternal appreciation to my advisor, Dr. Hong Koo Kim. It has been an honor and privilege for me to work with such a talented and pleasant scholar. I am immensely grateful to him for his thoughtful guidance, encouragement, and inspiration throughout my doctoral program. Also I express a sincere gratitude to the members of my dissertation committee, Drs. Joel Falk, Mahmoud El Nokali, Guangyong Li and Rob Coalson, for their taking time to serve on my committee and giving me invaluable comments and suggestions on this dissertation. I would like to thank my colleague, Yun Suk Jung, for his experimental support in this work. Jeff Wuenschell – Dr. Kim's former master student - also deserves recognition since his research provides helpful references to this study. I would like to thank my friends for helping me in my research and life.

I would like to thank my girlfriend, Dongxiao Li, for her love and support during my doctoral period. She kept encouraging and helping me whenever I encountered difficulties and frustration. Finally, I would like to thank my parents and sister for their love and great guidance in life.

1.0 INTRODUCTION

1.1 SURFACE PLASMONS

Surface plasmons, also called as surface plasmon polaritons (SPPs), are surface-bound electromagnetic waves that propagate along the metal/dielectric interface. By solving Maxwell's equations with appropriate boundary conditions, the SP dispersion relation can be obtained as follows [1]

$$k_{sp} = k_0 \sqrt{\frac{\varepsilon_M \varepsilon_D}{\varepsilon_M + \varepsilon_D}}. \quad (1)$$

From the equation, we can easily understand that for a dielectric constant of metal, $\mathrm{Re}(\varepsilon_M) < 0$ and $|\varepsilon_M| > \varepsilon_D$, $k_{sp} > k_0$. This wave vector relationship, i.e., the SP propagation constant's being greater than the free-space (dielectric side) wave vector results in a surface-bound wave that is evanescent in the normal direction to the metal surface, and is propagating along the interface without radiation. Because of the momentum mismatch, excitation of SPs by light requires compensation of momentum. Figure 1 shows the SPs fields and the dispersion relation, as well as the momentum (same as the wave vector, *k*) mismatch [2]. To provide the requisite momentum,

one needs a special structure that can generate an extra wave vector, such as the case of roughness of surface, e.g. gratings.

Metal corners can excite surface plasmons because the corner offers a continuum of spatial wave vectors and an appropriate extra wave vector needed for phase matching can be provided by the corner structure. Study of a single corner structure of a bulk metal with semi-infinite extension of both surfaces is therefore essential to understand the nature of SP excitation. In Fig. 2, a planar wave is incident along the vertical wall direction. How does the incident light interact with the corner or the corresponding surfaces? The corner can diffract the incident light and excite SPs at the same time. The scattering area will then be filled with the incident, diffracted and surface-bound waves, and perhaps even other wave components may need to be taken into account. In this study, we have employed the finite-difference-time-domain (FDTD) method in order to analyze the interaction of light with the structure. To analyze the contributions from individual wave components, it is necessary to extract each wave from the mixed waves. SPs are one of the major players and can be suppressed by adjusting the dielectric constant of metals such that they attenuate quickly. Another intrinsic difficulty in FDTD stems from the extremely short penetration depth (skin depth) in the metals supporting the high suppression of SPs. This requires the grid size be smaller than the penetration depth. Use of small grid sizes could result in unacceptably large computational complexity, especially for the case of large simulation windows. In this thesis we address these issues as described in the following sections.

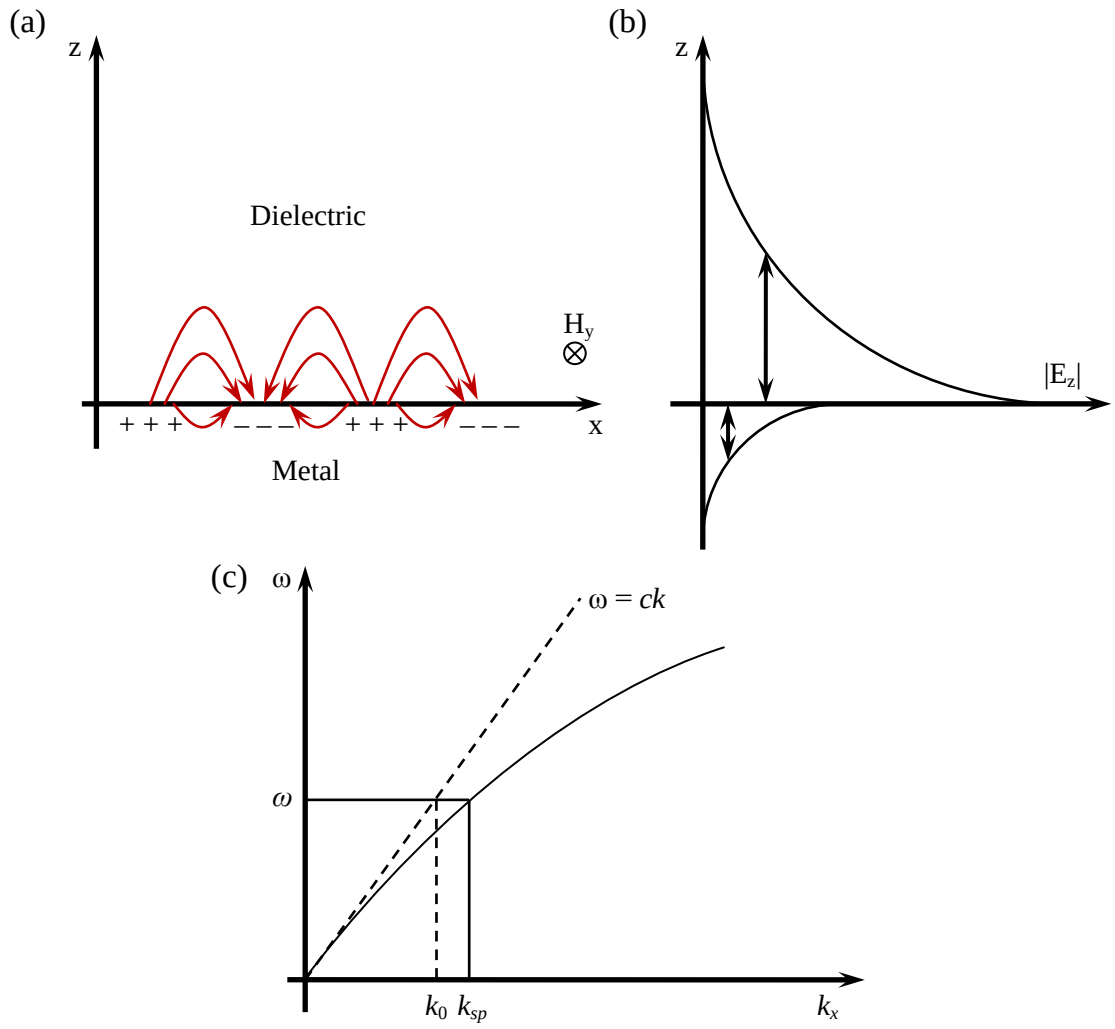


Figure 1. Surface plasmons. (a) SPs electromagnetic waves and surface charges at the interface between a metal and a dielectric material. It is noted that the transverse magnetic (TM) waves are required to excite SPs electrical fields normal to the surface to generate the surface charges. (b) The evanescent component of the field perpendicular to the surface exponentially decaying by distance away from the surface. Because of its evanescent property, SPs have the nature of non-radiation, which gives SPs bound and propagating upon the surface. (c) The dispersion relation of SPs in which the momentum mismatch (for the same frequency, the momentum of SPs is always larger than that of light in free space) stops the direct coupling between light in free space and SPs [2].

1.2 FDTD SUBGRID TECHNIQUE IN METAL SURFACE CALCULATION

In order to alleviate this numerical/computational problem, we have introduced a subgrid technique, which places subgrids, many smaller grid cells than primary grids, at specific locations in the computational window.

1.2.1 Errors Caused by Uniform Coarse Grids

To illustrate possible errors that may be caused by uniform coarse grid cells in FDTD calculation, we take a sample structure of a bulk metal with two corners (see Fig. 2). From the calculation (λ = 650 nm and ε = -5+655i) with uniform coarse grid cells FDTD algorithm, it is shown that the *y* component of electric field at left side and right side boundaries are asymmetric (Fig. 3). This error is attributed to the fact that the shifted Yee lattice results in choosing the values of H_z (*z* component of **H** fields) at (i+1/2, j+1/2) changing to at (i, j) (these two points have the same material properties, e.g. ε and μ), and therefore the metal boundaries divide the grid cells differently. This problem causes the difference of E_y (*y* component of **E** fields) at the boundaries of the metal.

This numerical error originates from the approximation involved in translating the differential form to the difference form. The finite size of the Yee cells results in the situation that the contour paths at the left side and the right side of the metal boundaries are not symmetric (Fig. 4). For the TE mode (electric field parallel to the *z*-direction case), H_y on the contour paths are only related to the permeability, which is nonmagnetic unity in metals. Therefore, the asymmetry of contour paths on both boundaries doesn't cause the asymmetry of H_y. However, for the TM mode case, E_y on the contour paths are asymmetric on both boundaries since they are

related to the permittivity different from 1 (the free space), and thus the asymmetric results of E_y arise (Fig. 3).

To confirm this conclusion, we can simply divide the permittivity into two different quantities in *x* and *y* directions. Now we only consider the Yee lattice with the values of H_z at (i+1/2, j+1/2). For the *x*-direction, we also keep the previous rules, ε_x (i, j) = ε_x (i+1/2, j+1/2). While, for the *y*-direction, we can set ε_y (i, j) = ε_y (i-1/2, j+1/2) at a certain boundary (e.g. left side or right side), thus metal boundaries cut Yee cells differently. Figure 5 shows the metal boundaries in the Yee cells after treating the left side or right side permittivity separately. Figure 6 shows the E_y distributions calculated by this treatment.

Actually, the difference of contour paths (or material properties inside) at both sides of the metal becomes much smaller when we reduce the size of Yee cells. If the grid size decreases, ε_y (i, j) on the contour paths boundaries are much closer to metal boundaries, therefore, ε_y (i, j) becomes more accurate in describing the permittivity values of metal boundaries. If we choose the grid size of 2.5 nm instead of 10 nm as we set previously, we can see the difference of E_y at both boundaries becomes much smaller as shown in Fig. 7.

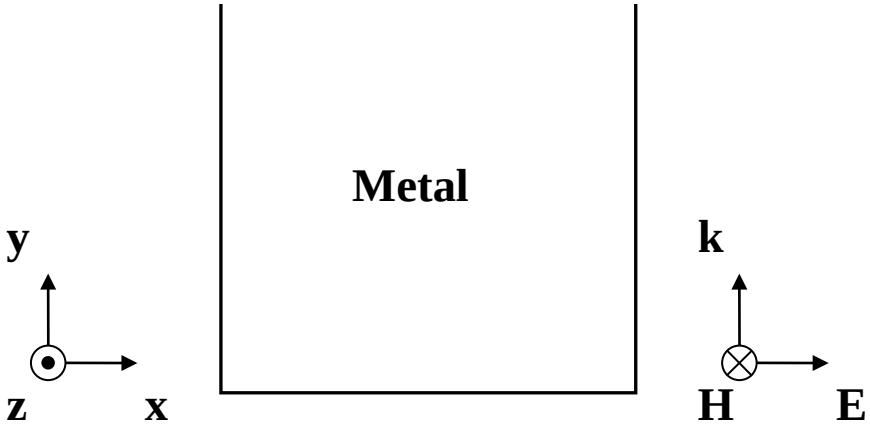
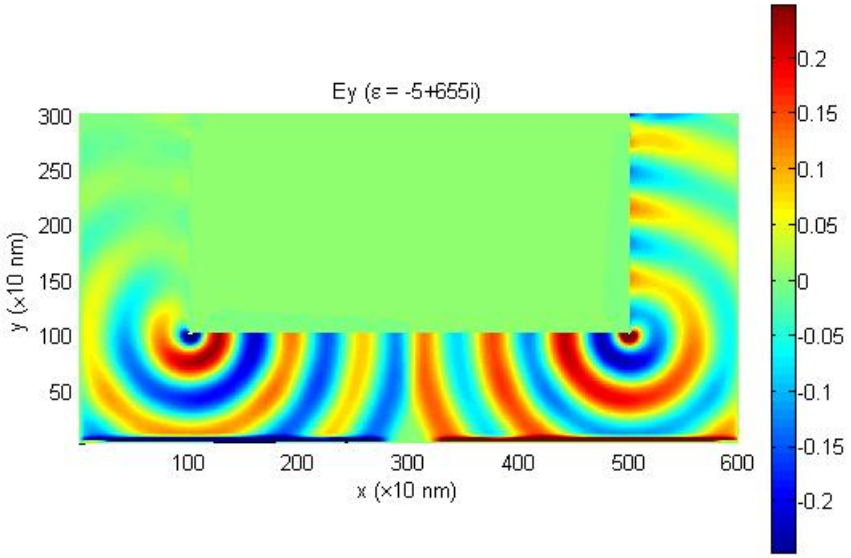


Figure 2. A bulk metal with two corners. A TM mode (**H** field parallel to the *z*-direction) light is incident to the *y*-direction.

(a)



(b)

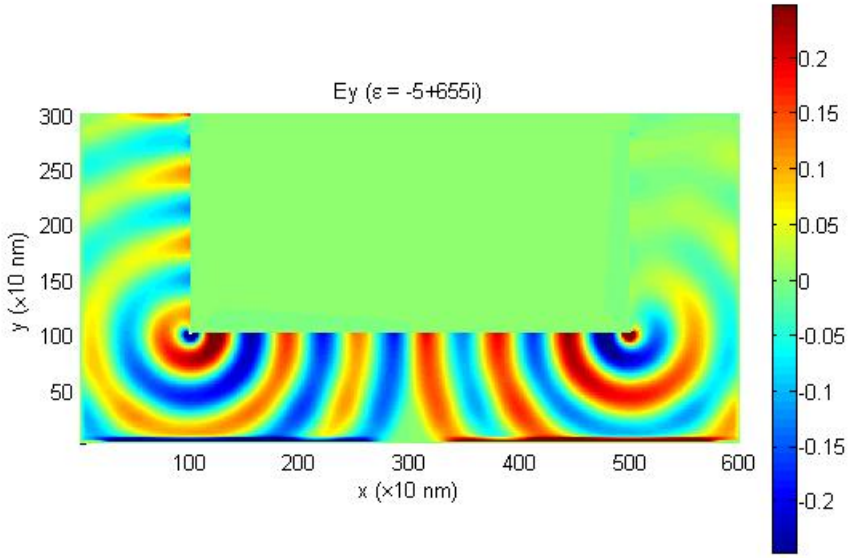
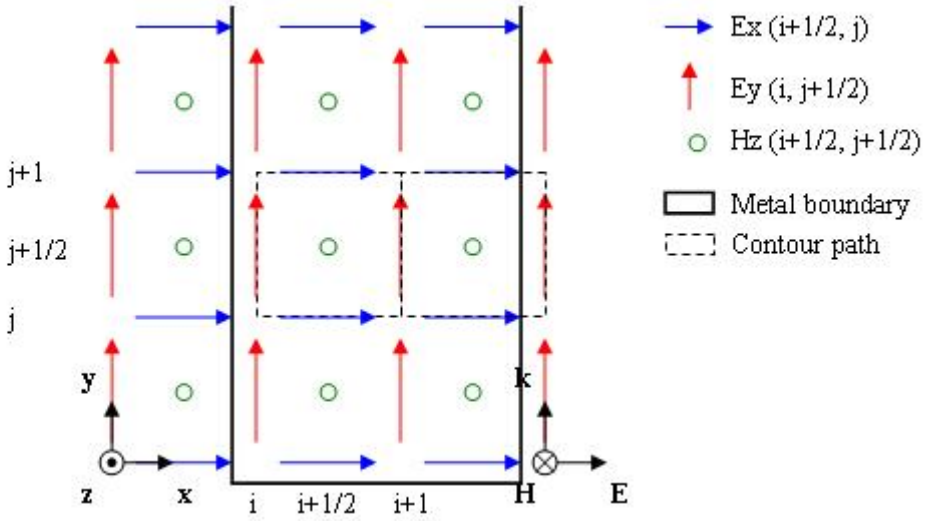


Figure 3. E_y distributions around a bulk metal. H_z is designed at (i+1/2, j+1/2) (a) and at (i, j) (b) in the Yee space lattice of the bulk metal (ε = -5+655*i* at 650 nm) with TM polarized incident wave. Grid size is 10 nm.

(a)



(b)

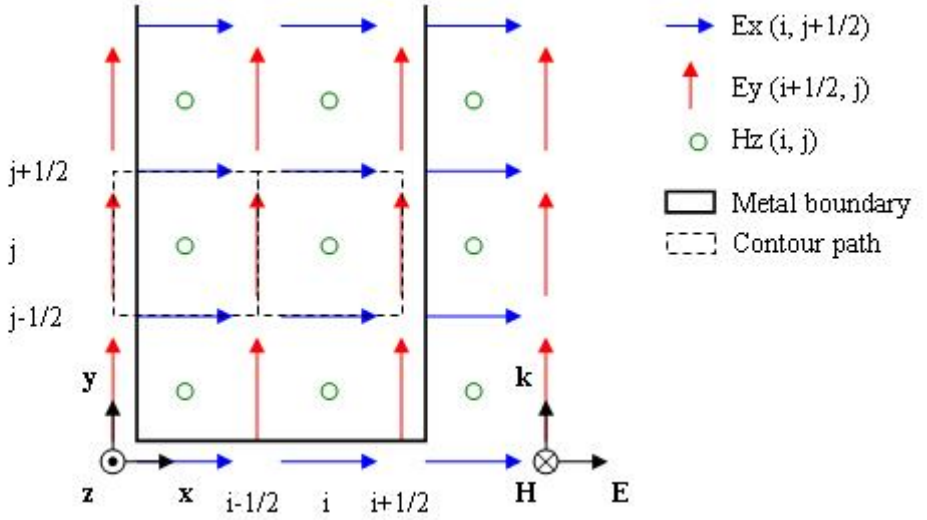
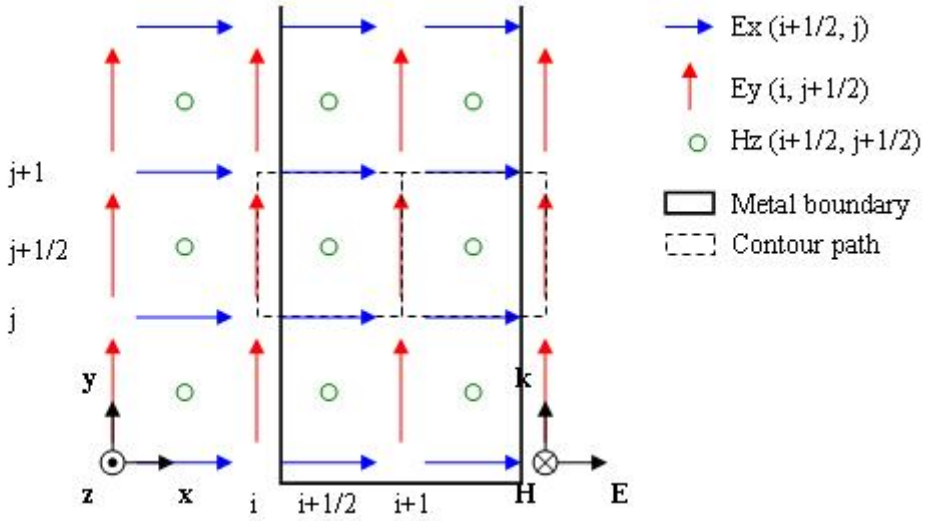


Figure 4. TM polarized Yee cells with H_z at different lattice points. H_z is designed at (i+1/2, j+1/2) (a) and at (i, j) (b) in the space lattice. Lattice points (i, j) and (i+1/2, j+1/2) will be treated as the same lattice, such as ε (i, j) = ε (i+1/2, j+1/2).

(a)



(b)

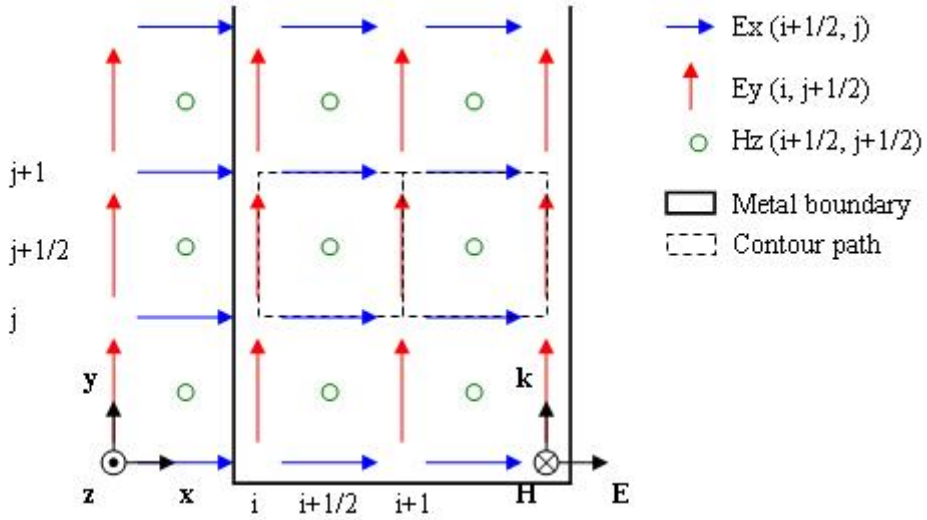
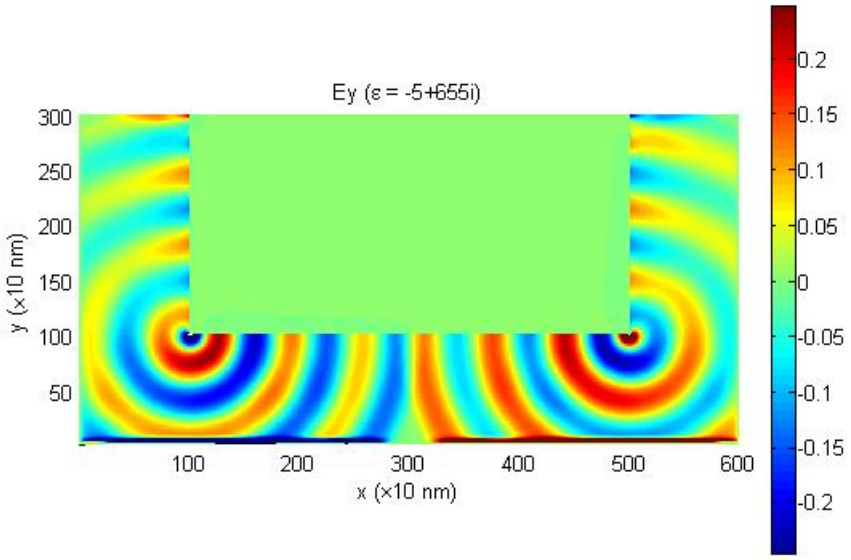


Figure 5. TM polarized Yee cells with H_z at (i+1/2, j+1/2) in the space lattice. Lattice points (i, j) and (i+1/2, j+1/2) are modified to be treated separately in the x and y directions, such as ε_x (i, j) = ε_x (i+1/2, j+1/2), ε_y (i, j) = ε_y (i-1/2, j+1/2) at left side (a) or right side (b) metal boundary.

(a)



(b)

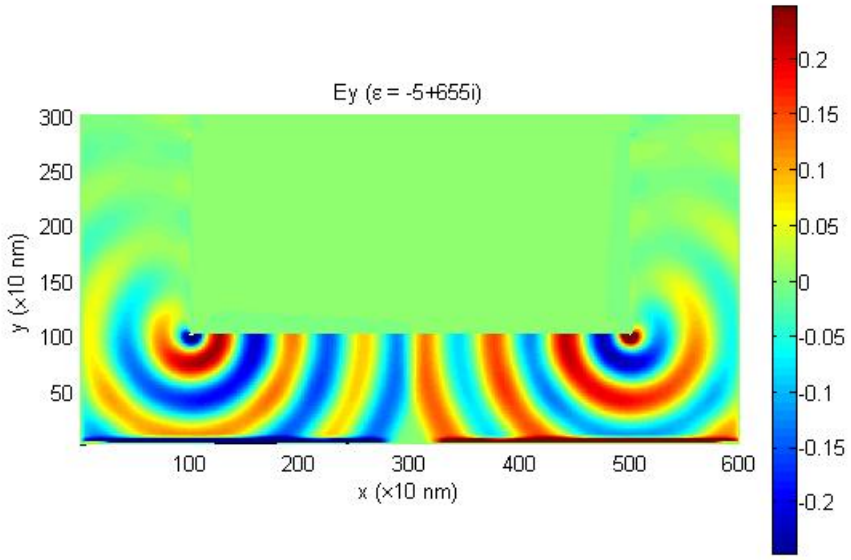


Figure 6. E_y distributions around a bulk metal with H_z at (i+1/2, j+1/2) in the modified Yee space lattice. The modification is designed at left side (a) or right side (b) metal boundary (depicted in Fig. 5) of the bulk metal (ε = -5+655*i* at 650 nm) with TM polarized incident wave. Grid size is 10 nm.

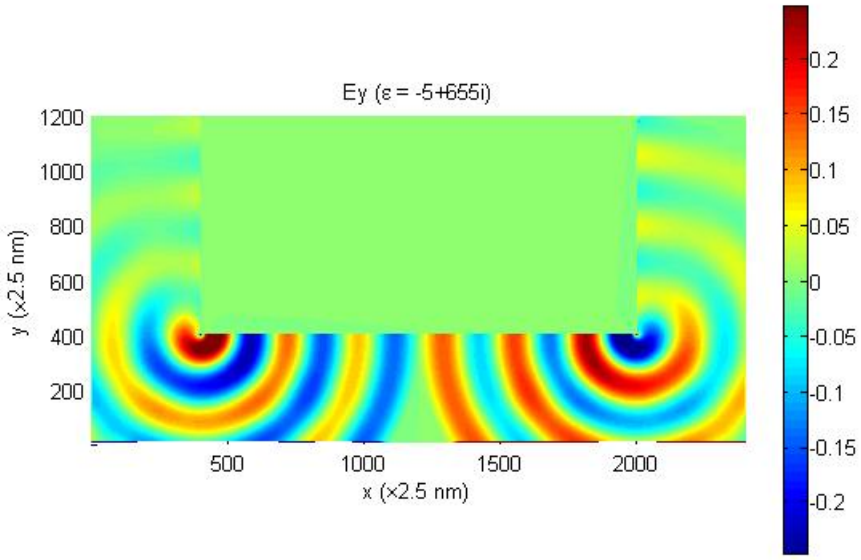


Figure 7. E_y distributions around a bulk metal with ultrafine grids. The bulk metal (ε = -5+655i at 650 nm) in the Yee space lattice is depicted in Fig. 4 (a) with TM polarized incident wave. Grid size is 2.5 nm.

1.2.2 Subgrid Technique

By decreasing the grid size and thus refining the mesh resolution, we can reduce the staircasing errors observed in the case of coarse FDTD grids. However, it requires much larger memory and longer computational time. To keep the computational load at a manageable level, we can use a more general approach, subgrid technique [3]. Subgrids are many small grid cells placed at specific locations in the computational area instead of uniformly modifying the primary grids. By nesting subgrids into coarse primary grids, we can refine the spatial grid size with the factor of 2^n:1. This kind of nested subgridding scheme is based on the Okoniewski scheme [4], which provides high performance for numerical stability and low boundary reflections between primary grids and subgrids.

The basic idea of the Okoniewski scheme is listed in Table 1. *E* and *H* are designated as primary grids fields, while *e* and *h* are designated as subgrids fields.

The geometry of nested subgrids in primary grids is shown in Fig. 8. A quarter-shifted subgrid spatial position corresponding to the primary grids leads to the smoothest transition. This is because the magnetic fields in the primary girds superposing with the magnetic fields in the subgrids simplify and optimize the spatial and temporal interpolation (e.g. simple averaging). To approximate the tangential values of *e* and *h* at the boundaries of subgrids region, we can choose a quadratic-function interpolation in time. For spatial interpolation, cubic spline method is better than linear method, but this can still lead to the numerical instability and reflections at the boundaries of subgrids region because of the abrupt spatial and temporal transition from the primary grid to the subgrid. By adjusting the smoothing factor in a cubic smoothing spline, we can ensure the stability of the system. Practically, if we set the smoothing factor as 0.9, no

tendencies for instability have been found up to more than 3500 time-steps (the propagation length of the surface plasma on the metal surface is up to 15 μm) in our tests.

To test the performance of the subgrid technique, we choose the geometry of tested example depicted as Fig. 9. The permittivity of metal is assumed to be ε = -5+0.655i. In our test we set the coarse grid size to be 10 nm, then the 1st level nested subgrid size is 5 nm and the 2nd level nested subgrid size is 2.5 nm (the refinement factor of 2^n:1). From Fig. 10 we can clearly see E_y field calculated by the subgrids is matched well with the fine grids.

Use of a subgrid technique can greatly save the memory and time and therefore improve the performance of computation. If we use simple fine grids to refine the results, for example, using the half grid size uniform fine grids instead of the previous coarse grids, to calculate the same sized computational window, we have to spend $2^3 = 8$ times longer time than before due to the doubled spatial lengths in *x*, *y* directions and the doubled temporal length. In the case of the subgrid technique, the extra time is spent only on the subgrids calculation, which is local and specific on a small region of subgrids. Table 2 is the system requirement of different grids size and nested subgrids level tested under personal computer with Intel 2GHz CPU. The width of subgrids region sketched in Fig. 9 chosen here for our test is 30 primary grids.

Table 1. Subgridding algorithm processes

Step	**Process**	**Time**
1	Update E	n
2	Update H	n+1/2
3	Interpolate in space at $t = n$ and in time to calculate the missing tangential e at the boundary using the nearest E in required region	n-1/2
4	Interpolate in space at $t = n$+1/2 and in time to calculate the missing tangential h at the boundary using the nearest H in required region	n-1/4
5	Update e and store the boundary values using the interpolated values at $t = n$	n
6	Update h and store the boundary values using the interpolated values at $t = n$+1/2	n+1/4
7	Update e	n+1/2
8	Update h	n+3/4
9	Interpolate in time to calculate the tangential H using h at $t = n$+1/4 and $t = n$+3/4	n+1/2

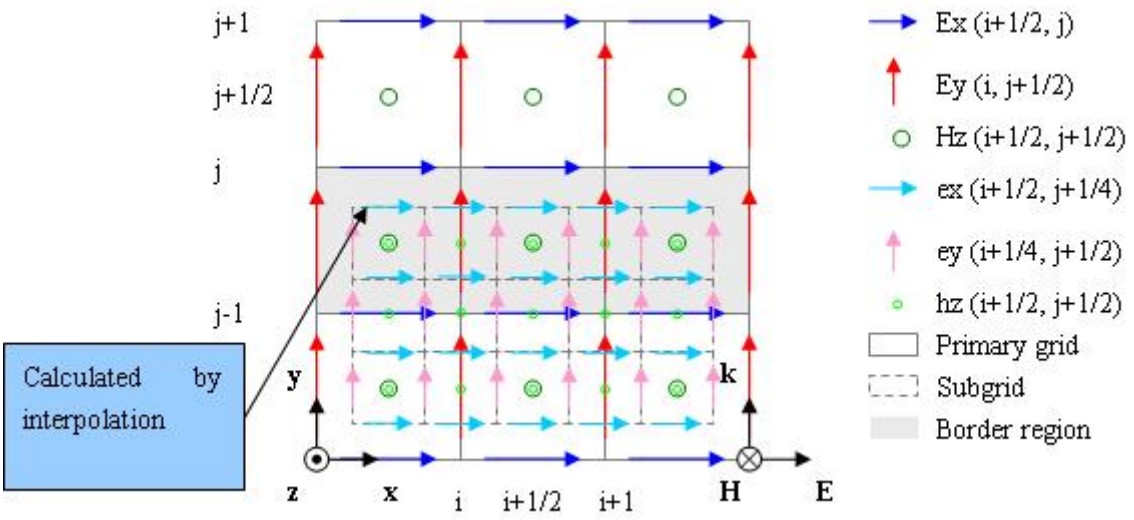


Figure 8. Sketch of subgrids geometry. At first, transfer data from primary grids boundaries to subgrids boundaries by interpolation, then calculate the subgrids data by Yee algorithm. At last, transfer data from subgrids to primary grids in the subgrids region.

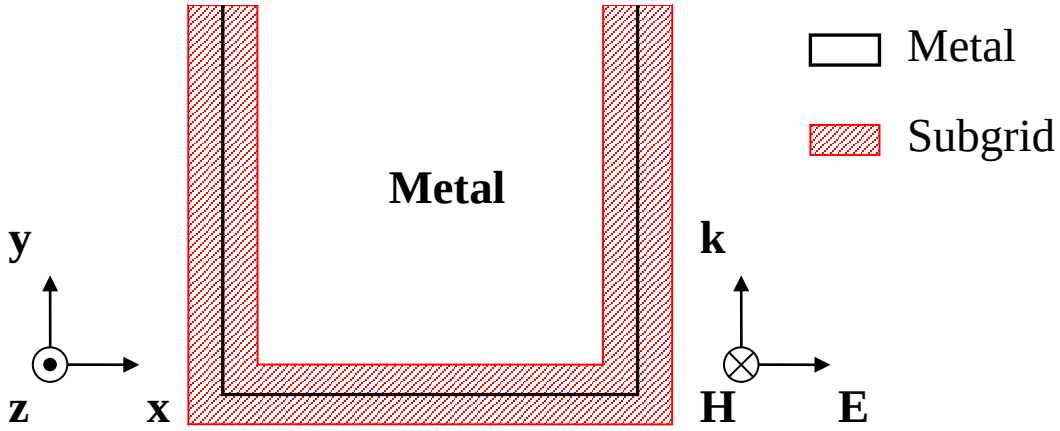


Figure 9. A bulk metal with two corners covered by subgrids area. Metal structure is described in Fig. 2. Subgrids area can occupy 1 level and 2 level nested subgrids.

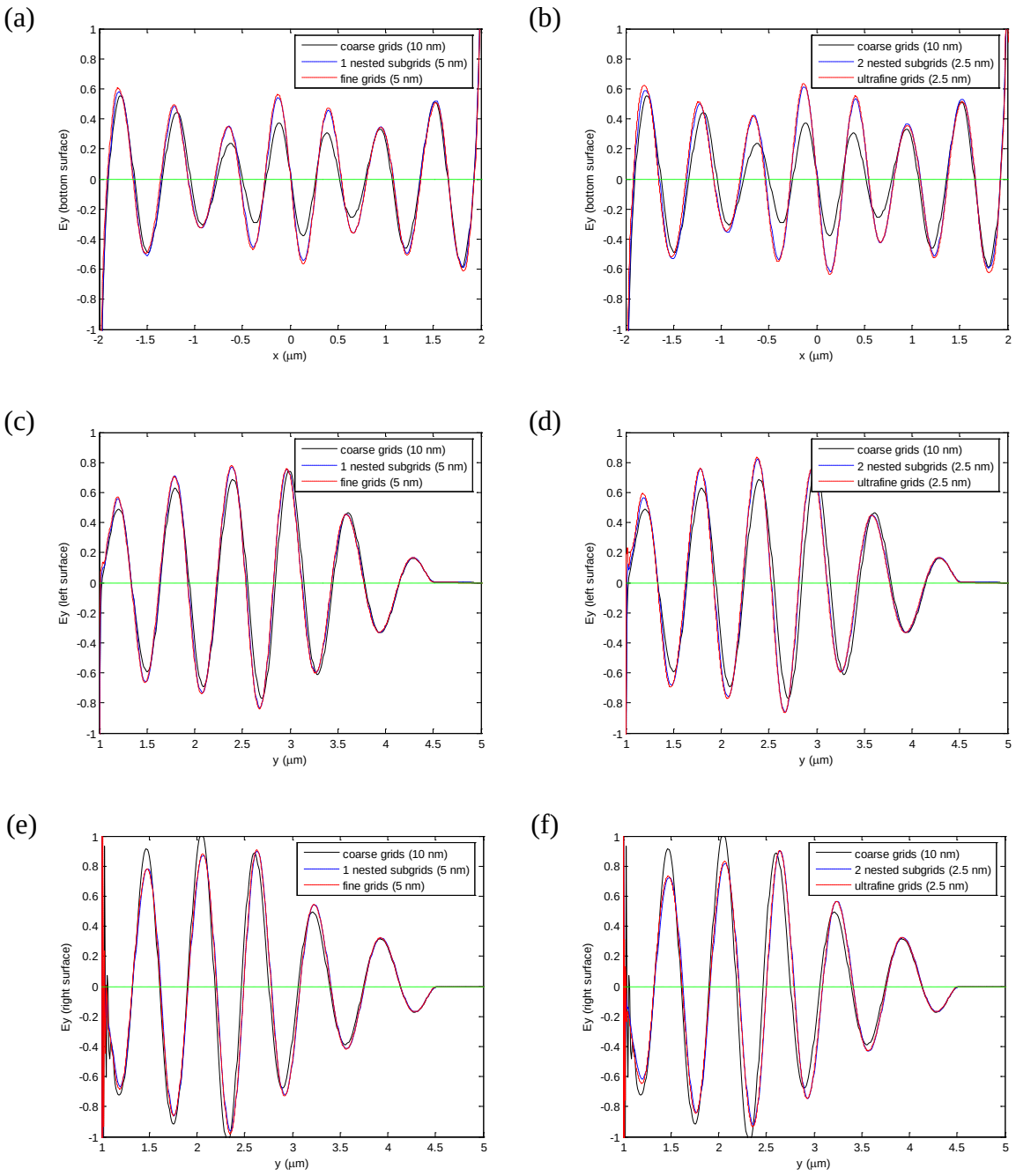


Figure 10. Scanned E_y along different surfaces with different grid sizes. E_y at bottom surface (a, b), left surface (c, d) and right surface (e, f) of a bulk metal (ε = -5+0.655i at 650 nm) with TM polarized incident wave calculated by coarse grids (grid size = 10 nm), fine grids (grid size = 5 nm) and 1st level nested subgrids (primary grid size = 10 nm, subgrid size = 5 nm) (a, c, e), ultrafine grids (grid size = 2.5 nm) and 2nd level nested subgrids (primary grid size = 10 nm, subgrid size = 2.5 nm) (b, d, f).

Table 2. System requirement for different nested subgrids levels and different grid sizes

Nested subgrids level & grid size	**Run time**	**Memory**
uniform coarse grids (grid size = 10 nm)	~8 min	~200 MB
1 level nested subgrids (primary grid size = 10 nm, subgrid size = 5 nm)	~10 min	~200 MB
2 level nested subgrids (primary grid size = 10 nm, subgrid size = 2.5 nm)	~40 min	~300 MB
uniform fine grids (grid size = 5 nm)	~1 hr	~450 MB
uniform ultrafine grids (grid size = 2.5 nm)	~9 hr	~1 GB

1.3 FDTD SIMULATION OF OBLIQUE INCIDENCE

The nanoaperture trasnsmission phenomena discussed above assumed normal incidence of an input beam. In order to study the more general case of oblique incidence, we need some preparation in FDTD simulation. Generating a usable plane wave obliquely incident to a flat metal is not a straightforward task. We first introduce a formulation for the boundaries of an incident plane wave. A plane wave touching the calculation boundaries or absorbing boundaries may result in numerical dispersion, and to avoid this potential problem a total/scattered (T/S) field formulation can be adopted. As shown in Fig. 11, in the T/S field formulation, the boundaries are set as the scattered fields, and the inner window is set as the total fields with the relationship: $F_{total} = F_{sca} + F_{inc}$. Thus, the incident fields do not appear in the scattered field region. This reduces the reflection at the absorbing boundaries and can avoid numerical dispersion. Because of the linearity of Maxwell's difference equations used in FDTD, slightly modifying the equations with the formula above at the interface between total/scattered field regions can keep the program going correctly with knowing the incident fields.

Now one remaining issue is how to generate a good oblique incident beam for iteration of FDTD. It is easy to generate a 1D normal incident beam as a look-up table thus each value of incident fields at the interface of T/S regions can be found at this look-up table. To obtain an oblique incident beam, we can just tilt the look-up table, and determine every incident field at the interface of T/S regions by finding the same wavefront in the look-up table using the interpolation [3]. This approach is available for the scattered field region filled with air, or with dielectric if we consider the wavefront by introducing the refractive angle. The advantage of this method is that the incident beam will propagate keeping the same complete wavefronts (a

complete wavefront appear to the calculation window at the same time). However, it is not good for metals because metals are too complex to be described simply by refraction (e.g., SPs may also be generated for TM-polarization). It is necessary to consider materials such as dielectric and metals into the look-up table and solve the problem for the materials that extend into scattered field region and touch the boundaries.

If we assume there is a periodic look-up table that describes a flat dielectric or metal film (in fact, which is also periodic), we can generate an oblique incident beam by using "self-iteration", which means we just set an input plane wave pulse (H_{inc}) at the source position and all other required field components are obtained by the periodic condition: $F_r = F_l \exp(-ikx\sin\theta_i)$, $F_l = F_r \exp(ikx\sin\theta_i)$, where F_l and F_r are any electromagnetic fields at left and right edges of a distance of x, respectively. Therefore, the fields at the window boundaries also can be obtained. Because of the instability of self-iteration approach at the beginning of the iteration, the upper part of each wavefront is cut-off, so it requires a longer calculation time (larger time steps) to get steady. However, if time is too long, results show that at T/F boundaries cross the metal surfaces, the numerical divergence occurs. To suppress the error, a thick layer of PML boundary can be used.

Figure 12 shows the results of generated periodic look-up table in free space (Fig. 12(a)), hitting on a flat silica (Fig. 12(c)) or on a Ag layer (Fig. 12(e)) with the incident angle of 45º. After subtracting the incident field at the incident side (below the silica or Ag layer, shown in Fig. 12(b)), a reflected field can be achieved (Fig. 12(d) for the silica case and Fig. 12(f) for the Ag case).

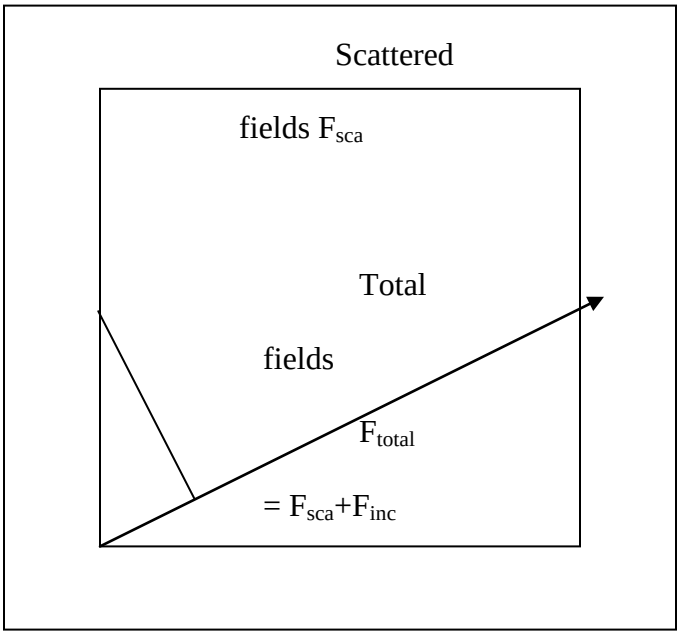


Figure 11. Schematic drawing of total/scattered field regions. The boundaries are set as the scattered fields, and the inner window is set as the total fields.

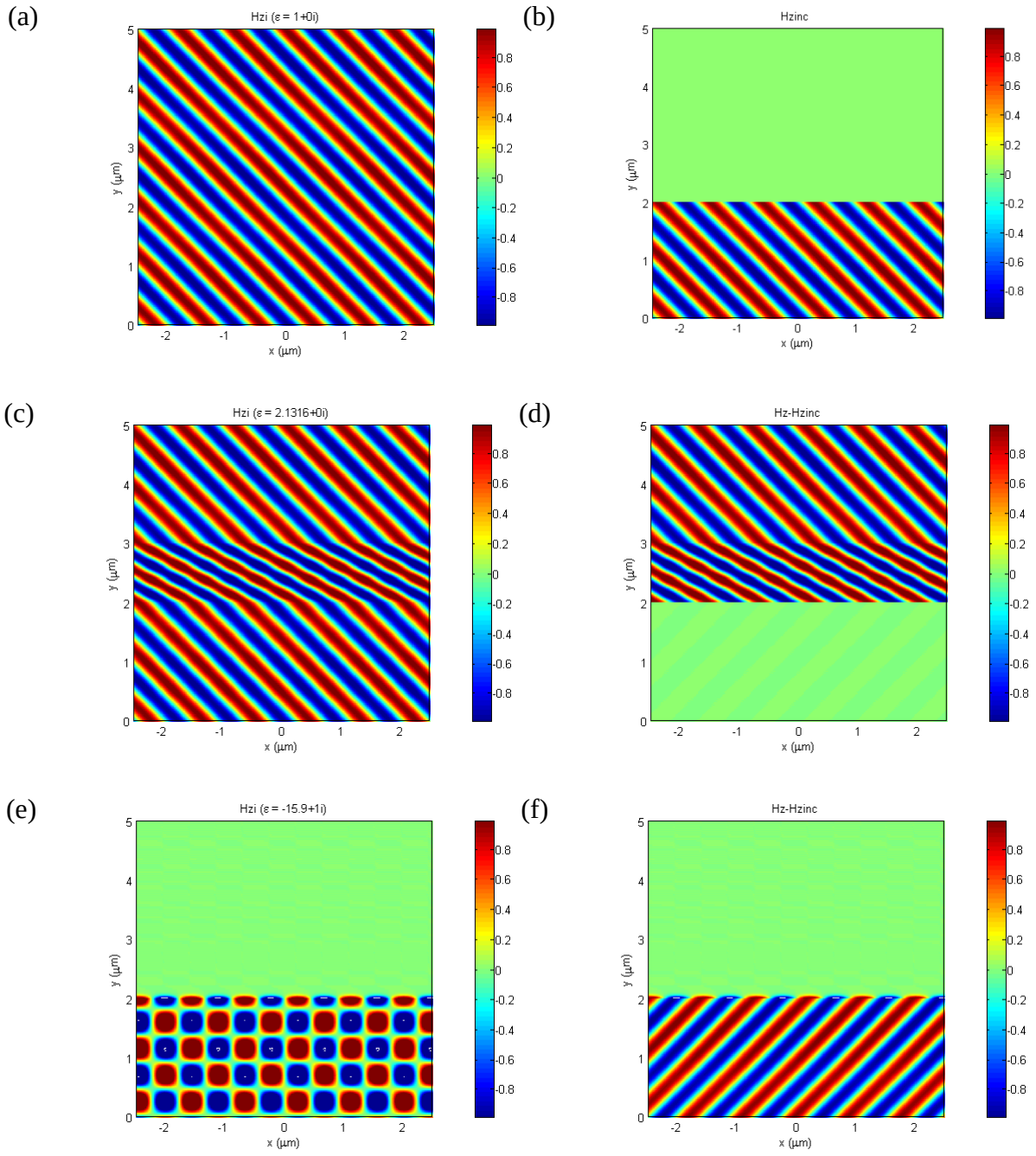


Figure 12. Generated periodic H_z look-up table for the oblique incident beam with the incident angle of 45º. The beam is in free space (a) and reacted with silica (ε = 2.1316) (c) and Ag (ε = -15.9+1.07i at 633 nm) (e) located at y: [2,3] μm. By subtracting the incident field in the region of [0,2] μm (b), the reflected fields by silica (d) and Ag (f) can be seen.

2.0 ANALYSIS OF INTERACTION OF LIGHT WITH A METAL CORNER

Based on the subgrid technique introduced in the previous section, we have performed FDTD simulation of a single metal wedge structure (the left corner in Fig. 9; the right part is replaced by absorbing boundaries). When a light wave hits a metal wedge structure, the metal surfaces respond to the incident light by generating both free-space and surface-bound waves. Here we present a physical model that elucidates electromagnetic interactions of an incoming planar wave with a simple semi-infinite 90^o metal wedge. We show that a metal wedge structure possesses an intrinsic capability of directing the incident power around the corner into the forward direction. Interplay of the boundary diffraction wave and the incident and reflection waves in the near field region of a metal corner is found to form a basis of the funneling phenomena that are commonly observed in metal nanoslit structures. Theory and experiment reveal that the incident wave propagating parallel to the sidewall destructively interferes with the boundary diffraction wave forming a depleted-energy-flow region along the glancing angle direction. A physical understanding of various electromagnetic phenomena associated with a metal wedge structure confirms rich potential of the simple structure as an elemental building block of complex metal nanostructures.

2.1 ROLE OF DIFFRACTION IN SHAPING ENERGY FLOW

It is well known that a metal wedge structure plays an essential role in scattering/diffracting light into free space and in exciting plasmon waves on metal surfaces [1,5,6,7,8,9,10]. Possessing a singular point at the corner, a wedge structure offers a broad-range continuum of spatial frequencies. This can help compensate the momentum mismatch between surface-bound plasmons and free-space waves, enabling interactions among the dissimilar wave components. The physical picture of the electromagnetic interaction with a metal wedge, however, is not clearly established, despite evident connection of the subject phenomenon to many related structures. A metal nanoslit, for example, has been commonly viewed as a basic building block of many plasmonic structures [2,11,12,13]. The funneling (field concentration) effect that is observable in individual nanoslits is considered one of the crucial factors in extraordinary transmission of nanoslit arrays [14,15,16]. In this work we consider a semi-infinite 90^{o} metallic wedge as a more elemental structure in understanding various phenomena associated with nanoslits. We show that the essential features of optical interaction with a nanoslit can be derived from the understanding of a metal wedge structure. This investigation focuses on the fundamental interactions that shape the energy flow distribution around the corner.

2.1.1 A model for boundary diffraction

In this work we have employed a finite-difference time-domain (FDTD) method in analyzing the interaction of light with a metal wedge. The electromagnetic field distributions show rather complicated patterns, and in interpreting the numerical simulation results we take the following approach. The waves produced by a metal wedge in response to an incident wave are categorized

into free-space or surface-bound waves (Fig. 13(a)). The free-space waves are further classified into reflection (with planar wavefronts) or diffraction (with non-planar wavefronts) components. Note that in this scheme the total fields are decomposed into planar and non-planar components by the geometry of wavefronts. This geometry-based classification of wave components is intended to help develop a physical picture of the complex electromagnetic phenomena that are analyzed by FDTD simulations.

Fig. 13(b), (c), (d) show a two-dimensional (2D) FDTD simulation result of a Ag wedge with a magnetic polarized planar wave. A planar incident wave was generated at the bottom side of the calculation window, and is expressed as $\mathbf{E}_{\mathbf{in}} = \hat{\mathbf{x}} E_0 e^{iky - i\omega t}$ and $\mathbf{H}_{\mathbf{in}} = (-)\hat{\mathbf{z}} H_0 e^{iky - i\omega t}$, where k is the propagation constant in free space. A perfectly matched layer (PML) was used for boundary condition. The dielectric constant of Ag was chosen as -17 + i1.15 at 650 nm wavelength [17]. A uniform grid size of 10 nm was used for the entire simulation window. (In some other cases that the dielectric function of metal is chosen such that the surface plasmon penetration depth in metal becomes significantly small (Fig. 18), the grid size of simulation needs to be adjusted to proportionally smaller ones. In that case, a sub-grid technique was used, with the primary grid size of 10 nm and the 2nd nested subgrid size of 2.5 nm [3].)

The snapshot images of the distributions of H_z- and E_x-fields (total fields) show the dominance of the incident and reflection wave components (Fig. 13(b), (c)). The diffraction or surface plasmon wave components, being of relatively minor intensity, are not clearly resolved in these images. In contrast, the E_y distribution can reveal the diffraction and surface plasmon waves (Fig. 13(d)). This is because the incident and reflection waves, being magnetic polarized, do not contain the E_y-field component. The concentric, circular-shaped wavefronts clearly indicate presence of a boundary (edge) diffraction wave in free space, while surface-bound

waves are visible on both perpendicular faces. In order to elucidate the diffraction component of the H_z and E_x fields, the incident and reflection components (planar wavefronts) are subtracted away one by one from the total field. First, upon the elimination of the incident wave, the reflection wave is clearly revealed, being present exclusively on the metal side, that is, in the 4th quadrant of the simulation window (Fig. 13(e)). The fine features of diffraction wave are buried under this strong reflection component, but further subtraction of the planar wavefronts of reflection uncovers circular wavefronts of diffraction (Fig. 13(f), (g)). (Here, for the subtraction of semi-infinite planar reflection wavefronts, first the infinite planar reflection wavefronts were calculated assuming an infinite metal surface with a planar wave normally incident to the surface. The right-half part ($x > 0$) of the planar reflection was taken and used as semi-infinite, uniform planar reflection wavefronts in the subtraction process. This simple geometry-based division/interpretation of the FDTD results implies that the non-uniform distribution of fields originating from the boundary diffraction is taken care of by the remaining, non-planar wavefront part of the total fields.)

The diffraction wave shows an antiphase distribution across the boundary between the planar reflection and circular diffraction regimes, that is, on the border of the 3rd and 4th quadrants. Note that the sharp discontinuity of diffraction wavefronts at the boundary stems from the step function profile assumed for the semi-infinite planar reflection wave. The amplitudes of the antiphase diffraction fields are found to be equal at the boundary and are 0.5 times the planar reflection field (Fig. 14). Therefore the total field (reflection plus diffraction) becomes continuous at the border, satisfying the requisite boundary conditions. It is interesting to note that the field amplitude of the diffraction component is unattenuated along the vertical ($-y$) direction although the field distribution broadens in the transverse direction. The anisotropic nature of the

radiation pattern suggests that the sources of the diffraction wave are not necessarily confined to the corner point itself, because otherwise a circularly symmetric pattern would be expected from a line (point) source. The simulation result rather suggests that an extended distribution of source elements is responsible for the observed diffraction pattern. In order to elucidate the nature of this boundary diffraction wave and its connection to the reflection wave, the electromagnetic response of the horizontal surface is modelled as follows.

In order to develop a better understanding of how the boundary diffraction wave (with non-planar wavefronts) is generated in conjunction with reflection of a planar incident wave by a metal edge, a model H_z-source of uniform reflection field ($H_{z,refl}$) was assumed to be placed on the bottom face of a perfect electric conductor (PEC) wedge, and the resulting emanating waves were analyzed by FDTD (Fig. 15(a)). The wavefronts produced by this semi-infinite uniform H_z source (*i.e.*, a half-plane $H_{z,refl}$ source) well match the FDTD simulation result of a metal wedge in most regions (See the 3rd and 4th quadrants of Fig. 15(a), and compare with those of Fig. 13(e)). In the 2nd quadrant, a discrepancy appears: the circular wavefronts partially remain strong around the vertical surface (Fig. 13(e)). This local, strong field around the sidewall is caused by the interaction of the diffraction wave with the sidewall as will be discussed below.

Placing a uniform H_z source on a metal edge is equivalent to assuming a uniform surface current on the metal surface. The implication of employing this model source can be understood as follows. For a planar wave normally incident to a metal surface, the metal responds by inducing a surface current, $\mathbf{J} = \mathbf{n} \times \mathbf{H}$. Here $\mathbf{H}$ refers to the total magnetic field vector on metal surface and $\mathbf{n}$ is the unit normal vector to that face. The induction of a surface current results in generation of a reflection wave, whose electric and magnetic fields are set such that they satisfy the boundary conditions on the metal surface. The reflection wave generated at the metal surface

(y = 0) is related to the incident wave as follows: $E_{x,refl} = rE_{x,in}$ and $H_{z,refl} = -rH_{z,in}$. Here r is the reflection coefficient of air/metal interface for normal incidence. For most metals $|r| \sim 1$, and the phase angle of r is close to 180 degrees, *e.g.*, 180 or 153 degrees for perfect conductor or Ag at 650 nm, respectively. The surface current that corresponds to the planar reflection wavefronts can then be expressed as $J = (1-r)H_{z,in} \cong 2H_{z,in}$. (The electromagnetic interaction with a metal wedge produces a complex distribution of fields. The induced surface current also takes a non-uniform distribution around the corner [18,19]. For the exact fields, the edge diffraction can be evaluated by taking the physical optics approach with a modified surface current [19].)

According to the geometry-based classification scheme taken in this work, the total magnetic field can be decomposed into two parts, one corresponding to the planar reflection wave and the other one to non-planar wavefronts (diffraction and surface-bound waves). The uniform surface current component by the planar reflection field is separated out from the total surface current, and is used as a model source in studying the response of the metal edge to an incident wave. As shown in Fig. 15(a), it is found that the half-plane uniform H_z source placed on the front surface of a metal wedge produces the non-planar diffraction wavefronts as well as the planar reflection wavefronts.

When a planar incident wave hits the bottom surface of a metal wedge, it produces planar reflection and non-planar diffraction (and surface plasmons) as discussed above. Note that these reflection/diffraction wave components propagate mostly in the 3rd and 4th quadrants of the simulation window. The 2nd quadrant region is affected by the presence of another diffraction wave component.

When an infinite-planar incident wave enters into the 2nd quadrant ($x < 0$ and $y > 0$), it suddenly becomes of semi-infinite extent confined to the air side of the space. Similar to the case

of an semi-infinite uniform H_z source placed on the bottom surface of the metal wedge for reflection, this propagating semi-infinite wave is expected to generate/bring non-planar diffraction components along with a planar one and to interact with the vertical sidewall of metal.

The nature of this boundary diffraction wave generation (by the propagating planar wave at the air side) and its interaction with the vertical sidewall of metal are further studied as follows. A half-plane uniform H_z soft (transparent) source is placed on the border of the 2nd and 3rd quadrants ($x < 0$, $y = 0$) in free space, and FDTD analysis is carried out for the emanating waves [Fig. 15(b), (c), (d) for the total H_z, E_y, and ($H_{z,total} - H_{z,pl}$) field distributions, respectively]. Comparing with the corresponding fields in Fig. 13(e), (d), (g), it is clear that the circular-shaped, anti-symmetric diffraction wavefronts originate from a semi-infinite uniform H_z source. It is important to note that unlike the case of reflection by the bottom surface, there is no surface current involved in describing the diffraction wave generation by a propagating semi-infinite planar wave in free space.

A semi-infinite planar source (both cases of Fig. 15(a) and 15(b)) is shown to produce composite, mixed-symmetry wavefronts, and this characteristic distribution of fields can be understood from the reasoning based on the Huygens-Fresnel principle: each point on a wavefront serves as a source of wavelets, and superposition of all the wavelets from the primary wavefront results in formation of next wavefronts [20].Imagine an infinite plane source of H_z in free space. The wavefronts emanating from the source would be perfectly planar without any circular fringe component. This is because the normal component of electric field (E_y) at any point would be cancelled out by anti-symmetric contributions originating from the opposing half-sections of the infinite plane source. As such the resulting field has only the horizontal

component (E_x). When the H_z source becomes of semi-infinite extent, cancellation of the normal component would not be perfect, especially around the edge region. The fringe (diffraction) field on the source-free side originates from the source side, that is, where the semi-infinite H_z is placed. By contrast, the fields on the source side lack contributions from the opposite side (source-free side).

Therefore the total field on the source side will be less the amount of the missing contribution than the otherwise perfect planar wavefronts. This symmetry-based reasoning provides an explanation on how a semi-infinite plane source produces such characteristic anti-symmetric circular wavefronts for H_z (and E_x) in conjunction with the semi-infinite planar wavefronts. We note that the field distribution of the waves (planar + circular) emanating from a half-plane source in free space takes the form of complementary error function, or equivalently of the Fresnel integral, an analytical expression well established in association with the knife-edge diffraction problem [21,22,23,24] (Fig. 16). [It is interesting to note in Fig. 15 that the total E_y field (Fig. 15(c)) shows an isotropic distribution whereas the total H_z field (Fig. 15(b)) is anisotropic. An analytical expression can be derived for the total fields originating from a half-plane uniform H_z source placed in free space. The half-plane source is considered as a semi-infinite array of densely-spaced line sources of infinite length (point sources when viewed on the x-y domain). The total fields are then calculated by integrating the contributions from all source elements. The expression for the E_y field is found to take the form of the complex complementary error function (or the complex Fresnel integral) and the field strength is a function of the distance from the edge (corner point) of the semi-infinite source. This functional dependence explains the isotropic nature (circular symmetry) of the E_y field.]

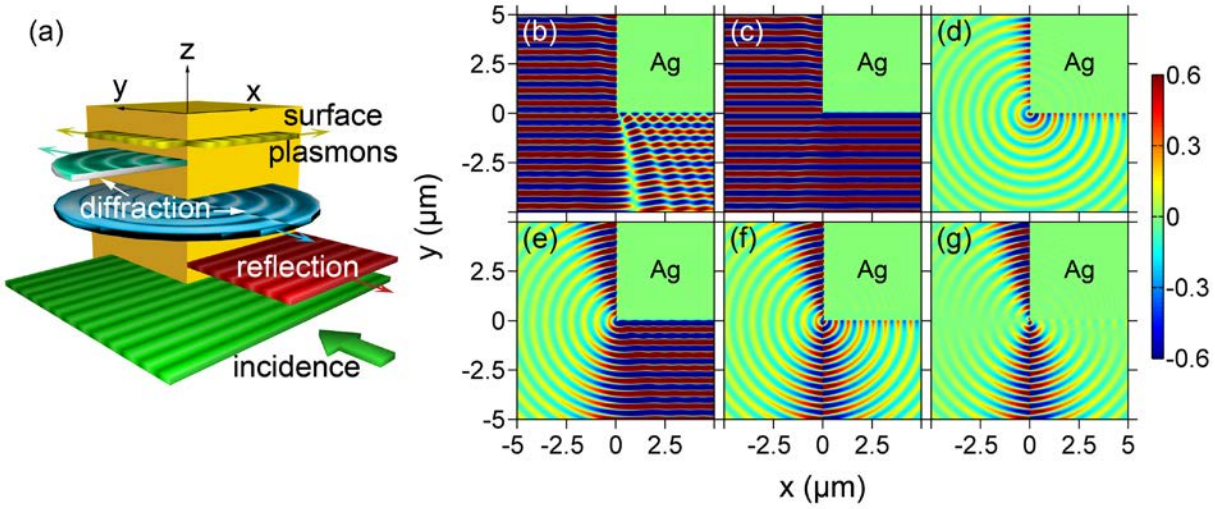


Figure 13. Electromagnetic interaction of a metal wedge. (a) Schematic illustration of free-space and surface-bound waves generated by a metal wedge for a normally incident planar wave: (from the bottom layer) incident; reflection; boundary diffraction; reflection of boundary diffraction; and surface plasmon waves. (b) to (g), Snapshot images of field distributions around a Ag corner calculated by FDTD for a 650 nm wavelength, magnetic polarized incident light: $H_{z,total}$ (b); $E_{x,total}$ (c); E_y (d); $H_{z,total} - H_{z,in}$ (e); $H_{z,total} - H_{z,in} - H_{z,refl}$ (f); $E_{x,total} - E_{x,in} - E_{x,refl}$ (g).

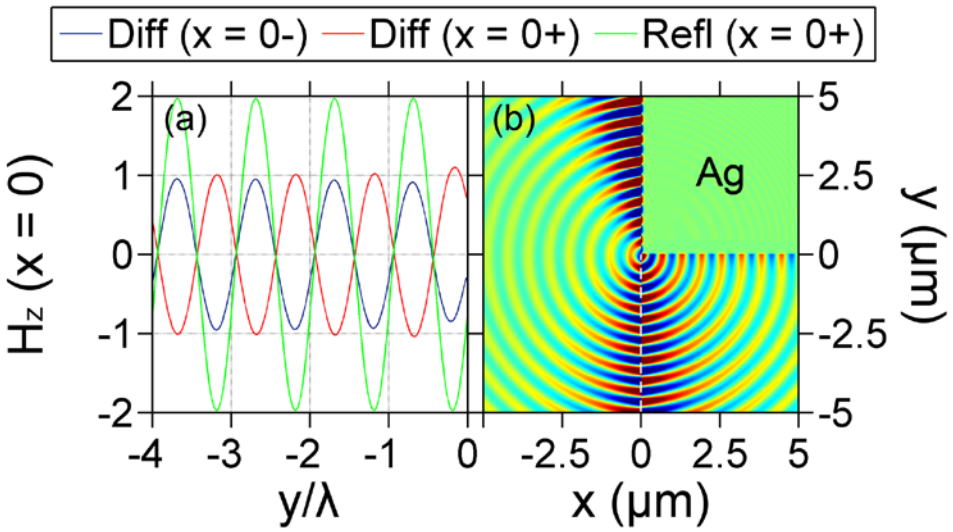


Figure 14. Anti-symmetric phase distribution of magnetic field (H_z) of a boundary diffraction wave around a Ag wedge. (a) H_z field profiles scanned along the y-direction at x = 0- (blue curve) or at x = 0+ (red curve). H_z field of planar reflection ($x \geq 0$, green curve) is also shown for comparison with boundary diffraction components in their phase and amplitude relations. (b) Snapshot image of H_z field distribution of a boundary diffraction wave. Note the opposite phase distribution across x = 0. A magnetic polarized light (650 nm wavelength) is incident from the bottom side of simulation window.

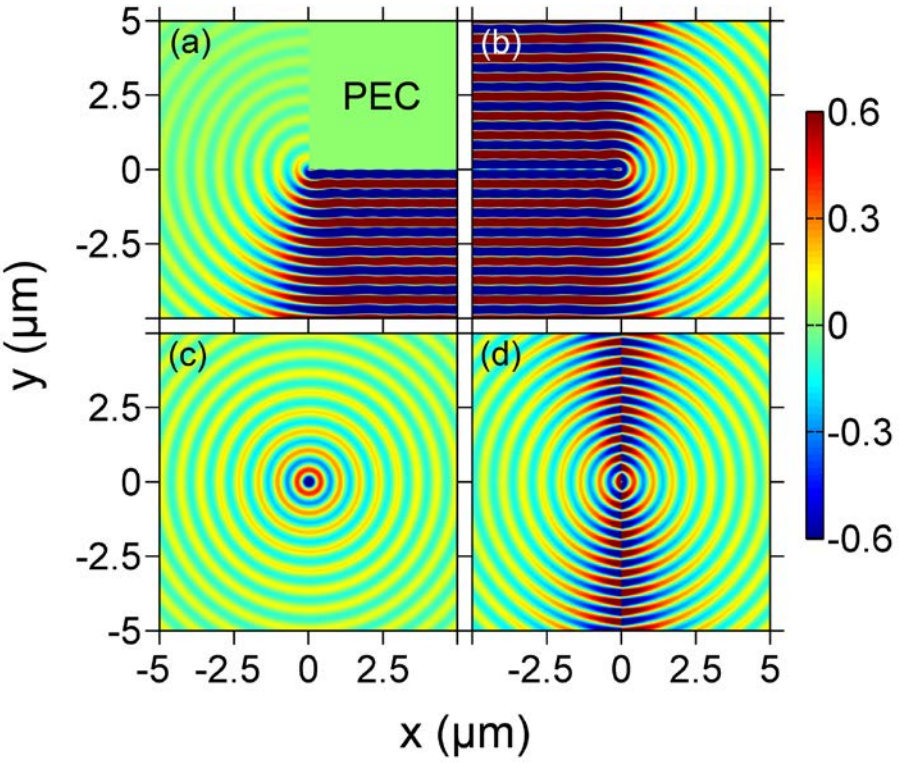


Figure 15. A model for boundary diffraction. (a) $H_{z,total}$ field distribution calculated by FDTD for a half-plane H_z source placed on the bottom face of a perfect electric conductor wedge. (b) $H_{z,total}$. (c) E_y. (d) ($H_{z,total} - H_{z,pl}$) field distributions calculated for a half-plane H_z source placed on the border of the 2nd and 3rd quadrants ($x < 0$, $y = 0$) in free space.

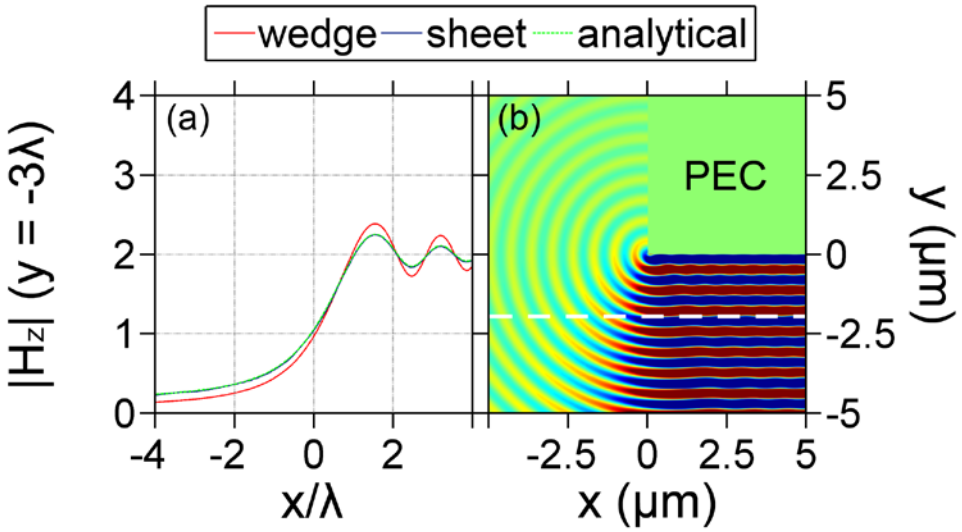


Figure 16. Field distribution of boundary diffraction waves. (a) $|H_z|$ distributions scanned along the x-direction at y = -3λ (λ = 650 nm): for a half-plane H_z source placed on the bottom face of a perfect electric conductor wedge (red); for a perfect electric conductor half-plane sheet placed at $x < 0$ and $y = 0$ (blue) with a magnetic polarized light normally incident to the PEC sheet. These FDTD calculated simulation results (red and blue) are compared with the analytical result (green) (after adjusting phases) calculated from the Sommerfeld half-plane problem, which concerns the same geometry as the PEC sheet case (blue) [5]. A good agreement is observed among the three results (especially between blue and green curves). (b) Snapshot image of H_z distribution for the case of a half-plane H_z source placed on the bottom face of a perfect electric conductor corner.

2.1.2 Interference of boundary diffraction: Field depletion and Funneling effect

Now we move on to the region around the vertical sidewall (the 2nd quadrant). The planar wavefronts of the total H_z field shown in Fig. 13(b) significantly bend in the region close to the sidewall. The ($H_{z,total} - H_{z,in} - H_{z,refl}$) and ($E_{x,total} - E_{x,in} - E_{x,refl}$) field distributions shown in Fig. 13(f), (g) also confirm strong presence of circular wavefronts localized in the same region. These seemingly peculiar distributions require a proper understanding. The incident wave entering into the 2nd quadrant can be modeled with a half-plane H_z source that is placed on the border with the 3rd quadrant ($x < 0$, $y = 0$). This H_z source is expected to develop fringe (diffraction) fields of the same nature as those of a half-plane H_z source placed in free space (Fig. 15(d)) discussed above. The boundary diffraction fields that are associated with the propagating semi-infinite planar wavefronts (Fig. 17(a)) and that fall onto the metal side (the 1st quadrant) are glancing incident in the region near/along the vertical sidewall and reflect back to the air side (the 2nd quadrant) (Fig. 17(b)).

The effect of this reflection at the vertical sidewall would be most significant along the glancing angle direction, because it corresponds to a shortest possible path that a wavelet can take in order to reach an observation point located near the vertical surface. For the case of an interface of a real metal with air, the reflection coefficient becomes -1 for glancing angle incidence (See the analysis below).

For a magnetic polarized light, the reflection coefficient of air/metal interface for oblique incidence is given as [6]:

$$r = \frac{\varepsilon_M \cos\theta_i - \sqrt{\varepsilon_M - \sin^2\theta_i}}{\varepsilon_M \cos\theta_i + \sqrt{\varepsilon_M - \sin^2\theta_i}}. \qquad (2)$$

For air/Ag interface at glancing angle incidence: ε_M = -17 + i1.15; $\theta_i \rightarrow 90^o$, $r \rightarrow \frac{-\sqrt{\varepsilon_M - 1}}{+\sqrt{\varepsilon_M - 1}} = -1.$

For air/PEC interface at glancing angle incidence: $\varepsilon_M \rightarrow i\infty$; $\theta_i \rightarrow 90^o$,

$$r \rightarrow \frac{\varepsilon_M \cos\theta_i - \sqrt{\varepsilon_M}}{\varepsilon_M \cos\theta_i + \sqrt{\varepsilon_M}} = \frac{\sqrt{\varepsilon_M}\cos\theta_i - 1}{\sqrt{\varepsilon_M}\cos\theta_i + 1}. \tag{3}$$

Since the reflection angles of boundary diffraction for glancing incidence are close to (but not equal to) 90º in reality, $\left|\sqrt{\varepsilon_M}\cos\theta_i\right| \rightarrow \infty$ and $r \rightarrow \frac{\sqrt{\varepsilon_M}\cos\theta_i}{\sqrt{\varepsilon_M}\cos\theta_i} = 1.$

The phase-inverted reflection component is then added to the direct propagation component on the air side (the 2nd quadrant). Since the original diffraction fields existing on the air/metal sides are antiphase to each other with equal amplitudes (blue and red circular curves in Fig. 17(a)), the reflection/superposition would result in (-1) times the original field of the planar component on the air side. (Recall that the diffraction field amplitude at the boundary is 0.5 times the original planar component.) When these total diffraction fields (dark blue circular curve in Fig. 17(b)) are added to the planar wavefronts on the air side (dark red straight line in Fig. 17(b)), a field depletion region develops near the sidewall. Note that the surface plasmons excited at the corner continue to propagate upward unaffected by the presence of the diffraction fields. For the particular case simulated in Fig. 13, that is, Ag at 650 nm wavelength, the surface plasmon propagation constant is similar to that of free-space wave.

The field depletion region that is expected to develop near the sidewall is not clearly visible in this simulation, because of the strong presence of surface plasmon fields in the adjacent region. When a different dielectric constant is assumed for metal such that the surface plasmon propagation constant significantly differs from the diffraction wave's, a field depletion region is clearly observed (Fig. 18). In the case of perfect electric conductor, the reflection coefficient at glancing angle incidence becomes +1, and reflection/superposition results in perfect cancellation of the antiphase diffraction fields on the air side. This implies the total fields near the sidewall would have only planar wavefronts without any local depletion of fields. The physical model of boundary diffraction discussed above provides a quantitative understanding of mirror imaging on metal surface [6].

We note that the field-depletion mechanism discussed above in conjunction with a metal corner can also explain the phenomenon commonly reported with metal nanoslit structures. It is well known that the radiation pattern of a nanoslit formed in plasmonic metal has a field depletion region in far fields along the glancing angle directions [12,21-23]. In contrast, for the case of a nanoslit formed in perfect electric conductor (or real metal operating in the longer wavelength range, in which the metal behaves like perfect conductor), the radiation pattern shows a uniform angular distribution over the entire span (π radian) without any depletion. The field-depletion in the case of a plamonic nanoslit can be understood as follows. As a light from a nanoslit radiates away (that is, as the radius of curvature of circular wavefronts increases), the wavefronts would look locally more planar, normal to metal surface. The interaction of the quasi-planar wavefronts with the metal surface in the glancing angle direction then becomes similar to that of semi-infinite planar wavefronts propagating along the sidewall of a metal

corner discussed above: incident fields are cancelled out near the metal surface via phase-inverted reflection of boundary diffraction waves.

Next we elucidate how the various wave components interplay in the near- to far-field regions, shaping the energy flow distribution around the corner. Fig. 19(a), (b), (c), (d) show the Poynting vector distributions ($\left|\langle S\rangle\right|$, $\langle S_x\rangle$, and $\langle S_y\rangle$) calculated for a metal (Ag or perfect electric conductor) wedge. Note that the energy flow in most regions is dominantly along the *y*-direction, that is, the $<S_y>$ component is an order of magnitude stronger than the $<S_x>$ component. The 2nd quadrant corresponds to the region where primarily the planar incident wave and the circular diffraction wave interplay in far fields. Note the contrasting difference between PEC and Ag cases (Fig. 18(a), (b)): the Ag wedge shows virtually-zero energy flow in the region along the sidewall whereas the PEC wedge does not reveal such depletion. This is due to the fact that in Ag case the phase-inverting reflection of a boundary diffraction wave into its anti-symmetric counterpart in the 2nd quadrant results in cancellation of incident fields along the side wall. In PEC case, the sidewall reflection does not induce any phase change and the anti-symmetric diffraction wavefronts cancel each other upon reflection. Therefore the incident wavefronts remain unaffected by reflection of boundary diffraction components. The energy flow carried by surface plasmons is clearly visible in the near field region along the sidewall of Ag. The quasi-parabolic fringe pattern observed in the intermediate- to far-field region indicates interference of two propagating waves, one with planar and the other with circular wavefronts [See Fig. 5 and reference 24]. In the 3rd quadrant, the planar incident and circular diffraction waves are counter-propagating, therefore interference of the two waves produces a fringe pattern, distinctly different from that in the 2nd quadrant. Because of the contra-directionality, the fringe spacing is significantly smaller than the co-propagating case. The 4th quadrant reveals much

complex patterns. This is because two extra wave components are present in this region, the reflection wave and the surface plasmons besides the incident and diffraction waves.

Now we consider how a metal wedge structure would possibly induce a funneling effect on a planar wave incident to the wedge. Figure 19(e) shows the Poynting vector ($\langle S_x \rangle$) distribution scanned along the border of the 3rd and 4th quadrants ($y < 0$; $x = 0$). The energy flow in the horizontal direction shows an oscillatory behavior: it starts with a relatively large negative value at the surface ($y = 0$) and then alternates between the positive and negative regimes with an oscillation period of $\lambda/2$. The amplitudes of the negative peaks are found to be persistently greater than the neighboring positive peaks. This indicates a net energy flow into the negative x-direction ($\int_{-\infty}^{0} \langle S_x \rangle dy < 0$), that is, a transfer of energy from the 4th quadrant to the 3rd quadrant crossing the border ($x = 0$; $y < 0$). Figure 19(f) shows the Poynting vector ($\langle S_y \rangle$) distribution scanned along the border of the 2nd and 3rd quadrants ($x < 0$; $y = 0$). The energy flow along the positive y-direction shows relatively symmetric undulation around the incident flux level, except for the region very near to the corner ($-0.1\lambda < x < 0$). The upward energy flow crossing the entire border ($x < 0$; $y = 0$) is found to be slightly smaller than the incident energy flow, that is, $\int_{-\infty}^{0} \langle S_y \rangle dx < \int_{-\infty}^{0} \langle S_y \rangle_{in} dx$. In a local view, however, the region near the corner exhibits a net positive flow of energy beyond the level of the planar incident flux (directed to the positive y-direction).

In reviewing the two energy flow distributions (the horizontal and vertical scans around the corner) it becomes evident that a funneling phenomenon does occur, but only in the near-field region around the corner. Note that in Fig. 19 the funneling region is not clearly resolved, but when magnified, the corner region reveals funneling (Fig. 20). The diffraction wave

component that is being produced in conjunction with reflection from the front surface causes undulation of energy flow. The near-field region around the corner, where the funneling phenomenon is confined to, is estimated to be approximately 50-nm and 80-nm wide on the front and sidewall surfaces, respectively.

In order to better understand the funneling process around the corner, the Poynting vector $\langle S_x \rangle$ is decomposed into two parts for the case of PEC: one contributed by boundary diffraction wave itself and the other by cross-coupling of boundary diffraction wave and planar incident and reflection waves.

$$\langle S_x \rangle = \frac{1}{2}\mathrm{Re}(E_y \times H_z^*) = \frac{1}{2}\mathrm{Re}(E_{y,diff} \times H_{z,diff}^*) + \frac{1}{2}\mathrm{Re}(E_{y,diff} \times H_{z,pl}^*) \equiv \langle S_x \rangle_{diff} + \langle S_x \rangle_{cross} \quad (4)$$

Here we note that the planar incident and reflection waves do not contain E_y component. The E_y fields are contributed by boundary diffraction waves and they converge to surface plasmon fields on metal surface in the case of a Ag corner. The amplitude and phase plots of the field components ($E_{y,diff}$, $H_{z,diff}$, and $H_{z,pl}$) reveal that the cross-coupling term $\langle S_x \rangle_{cross}$ is dominant over the self-diffraction term $\langle S_x \rangle_{diff}$ (Fig. 20(e), (f), (g), (h)). This is mainly due to the fact that the H_z field on the metal side is supplemented by an in-phase reflection wave and the total planar wave component becomes much stronger than the diffraction component of H_z. Considering the phase distributions of diffraction fields around the corner (symmetric E_y and anti-symmetric H_z), a diffraction wave itself would radiate the energy away from the corner. Interplay of boundary diffraction fields (E_y with symmetric, linearly rising phases) and planar wavefront fields (H_z with symmetric constant phase), however, enables funneling around the corner. Note that the

diffraction field E_y at $x = 0$ is almost phase-inverted compared to the planar wave field H_z. Due to this phase relationship and the linearly rising phase profile of E_y, cross-coupling of the two fields yields forward energy flow ($S_x < 0$ and $S_y > 0$) around the corner in the narrow region of front metal surface (Fig. 20: $x <$ ~100 nm). Note that inside the Ag metal corner Poynting vectors of significant strength appear with orientations opposite to that in the air side (Fig. 20(c): on the horizontal surface of the wedge). The tangential components of Poynting vectors associated with surface plasmons are opposite to each other inside and outside the metal surface, and this is due to the fact the dielectric constant of metal is largely negative real [25,26,27,28].Note that a PEC corner (Fig. 19(e), (f), Fig. 20(d)) shows the same or even slightly stronger funneling than the plasmonic metal case. This indicates that surface plasmons are not essential to induce a funneling phenomenon [29,30]. Overall this analysis elucidates that a metal corner possesses an intrinsic capability of funneling light, which is most commonly observed in a metal nanoslit structure. The energy flow distributions around a corner further suggest that when two corners are brought together to form a slit, the funneling effect of the slit would show characteristic dependence on slit width (Fig. 19(e)). In order to maximize the funneling effect (the transmission normalized by slit width), slit width needs to be smaller than ~160 nm. For larger widths, the amount of light coupled into a slit will show an oscillatory behavior (with a period equal to λ) as a function of slit width [31,32,33].

A silver wedge structure was formed by depositing a thick Ag layer on a cleaved edge structure of a GaAs wafer. A (001)-oriented GaAs wafer (356-μm-thick, single-side polished) was cleaved along the (110) cleavage plane by a razor blade. This process produces a 90-degree wedge structure defined by two crystal planes (001) and (110). The cleaved wafer (1 cm x 2 cm size) was introduced into a vacuum deposition chamber (base pressure of 10^{-6} Torr), and a 200-

nm-thick Ag layer was deposited on the edge facets by thermally evaporating Ag (4N8 purity: Alfa Aesar). The sample face, the (001) plane, was 45-degree tilted to the evaporation flux so that there be no shadow effect on the corner area and both facets receive the same amount of Ag flux during deposition.

Optical intensity profiles around the Ag corner structure were measured by scanning a nanoapertured probe (Veeco Aurora NSOM probe 1720-00: 100-nm-thick Al coated; 80-nm aperture diameter). A magnetic polarized He-Ne laser beam (633 nm wavelength, 1-mm beam diameter) was incident to the corner, normal to the (001) surface. The scanning probe was tilted 30 degree off from the vertical sidewall, the (110) plane, and was scanned from the corner (0 μm) to 6 μm up in the vertical (y) direction, and from the sidewall (0 μm) to -6 μm in the horizontal (x) direction. The step size of scan was 50 nm and 125 nm in the horizontal and vertical directions, respectively.

Figure 21 shows an experimental data of the energy flow distribution around the Ag corner measured by a scanning probe technique. The probe axis is tilted 30 degree off from the vertical direction (Fig. 21(a), (b)), and therefore the measured intensity more closely represents the vertical component of Poynting vector ($<S_y>$) than the horizontal component ($<S_x>$). The measured fringe pattern (Fig. 21(c), (d)) well matches the simulation result (Fig. 19(b), (d)).

A well-defined ‘dark’ (virtually-zero energy flow) region is observed near and along the sidewall. The sharply-rising intensity profile near the metal surface corresponds to the surface plasmon power propagating along the positive y-direction. The energy flux incident to this ‘dark’ region parts into the surface plasmon and free-space fluxes. The transverse components of Poynting vectors in this region take opposite signs, splitting the incident flux and diverting them away from the region, one toward the sidewall (with a positive value of $<S_x>$) and the other into

free space (with a negative value of $<S_x>$) (Fig. 19(c), (d)). The fringe locations are determined by interference of planar incident and circular diffraction wavefronts. The $<S_y>$ profile scanned at $y = 0$ along the negative x-direction shows a fringe spacing well-matching the free-space wavelength of light (633 nm) (Fig. 21(e), (f), (g)).

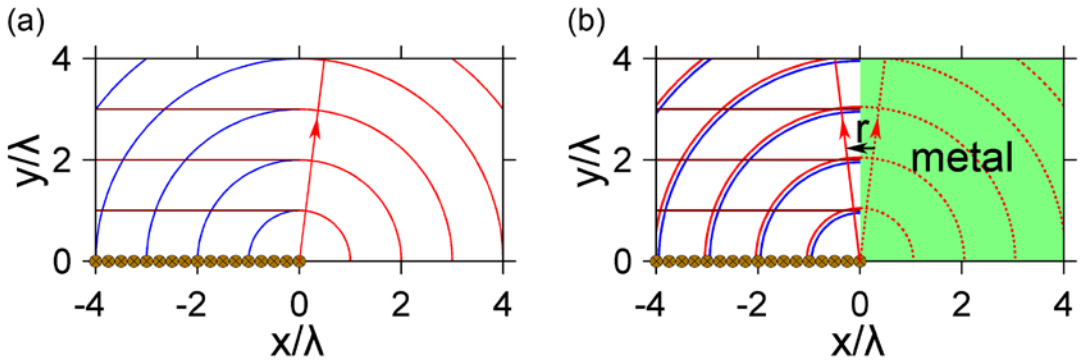


Figure 17. Schematics of boundary diffraction wavefront generation by a propagating semi-infinite planar H_z source in free space and its interaction with the vertical sidewall of a metal wedge. (a) A half-plane H_z source placed in free space (an array of brown dots with crosses inside: placed on the border of the 2nd and 3rd quadrants, $x < 0$, $y = 0$) generates anitiphase non-planar boundary diffraction wavefronts (red and blue solid curves) besides the planar wavefronts (brown solid lines). (b) When a metal wedge is introduced in the 1st quadrant (green block: $x > 0$, $y > 0$), the non-planar diffraction wave components in the 1st quadrant (red dotted curves) reflect back to the 2nd quadrant (red solid curves) and are superposed to the diffraction components propagating in that region (blue solid curves). The 'r' indicates the reflection coefficient at the vertical sidewall. The red arrows denote the propagation direction of wavefronts.

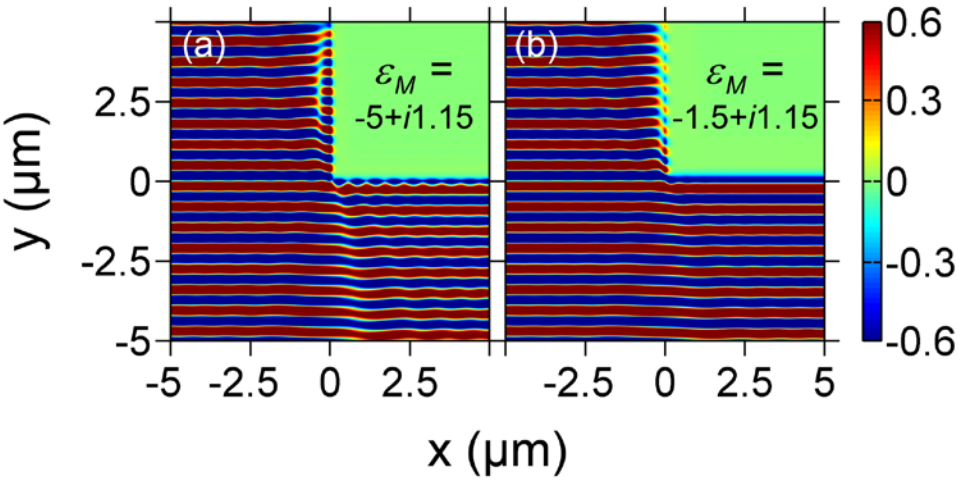


Figure 18. Snapshot images of magnetic field (H_z) distributions around a metal wedge. Different dielectric constant values are assumed for FDTD calculation: (a) ε_M = -5 + i1.15; (b) ε_M = -1.5 + i1.15. A magnetic polarized light (650 nm wavelength) is normally incident from the bottom side of the simulation window. Presence of field-depletion region is clearly observed near the vertical sidewall of metal, although hindered by coexistence of surface plasmon fields.

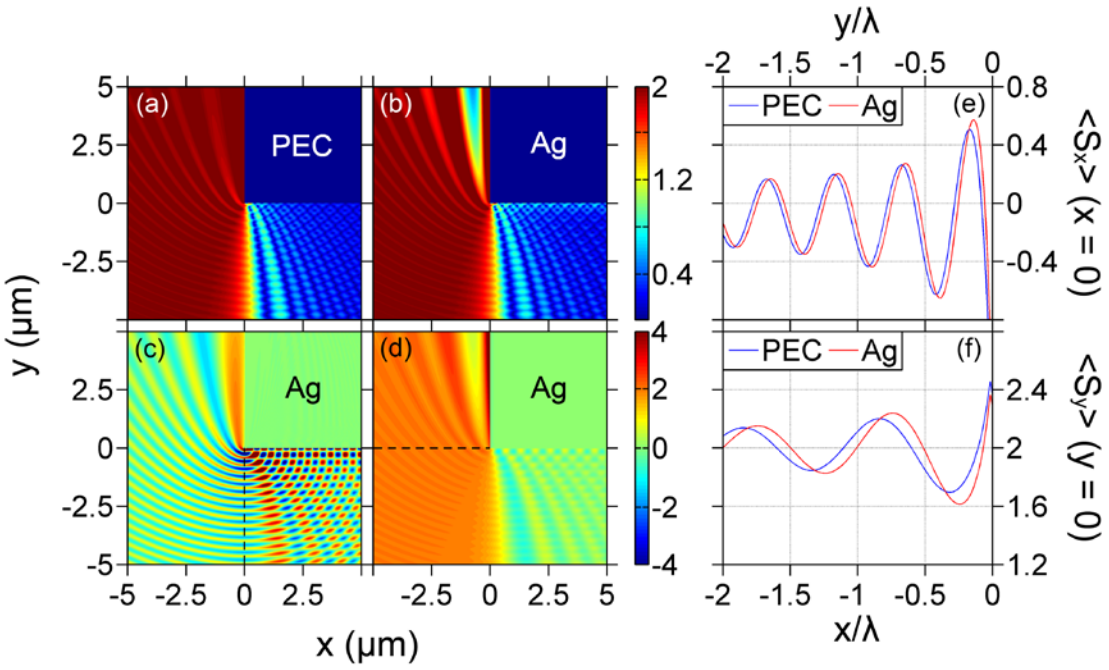


Figure 19. Energy flow distributions calculated by FDTD. (a) $|\langle S \rangle|$ for a perfect electric conductor wedge. (b) $|\langle S \rangle|$ for a Ag corner. (c) $\langle S_x \rangle$ (10 times magnified) for a Ag corner. (d) $\langle S_y \rangle$ for a Ag corner. (e) $\langle S_x \rangle$ scanned along the negative y-direction at $x = 0$. (f) $\langle S_y \rangle$ scanned along the negative x-direction at $y = 0$.

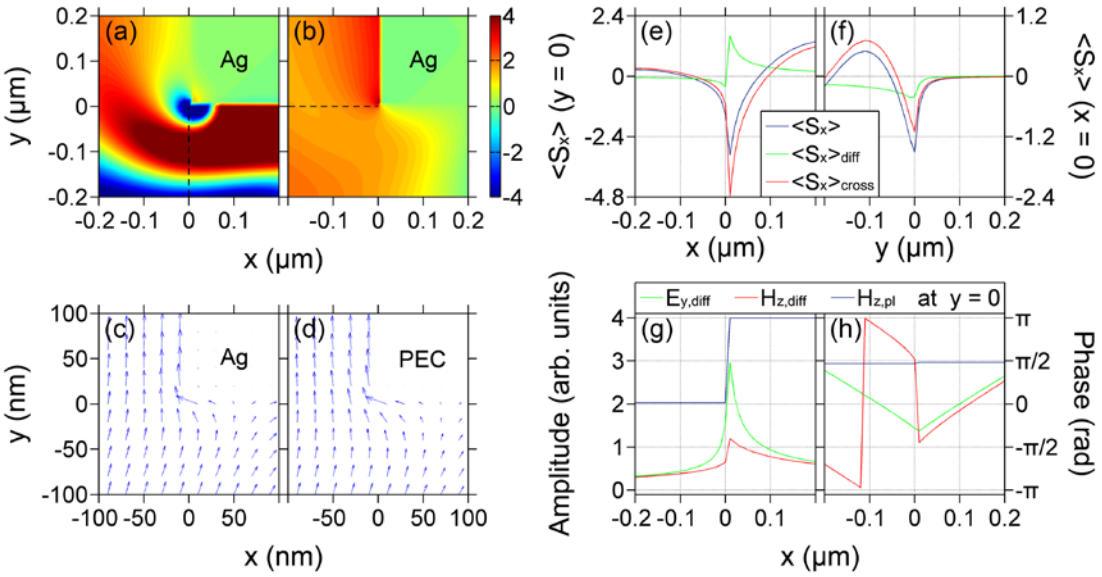


Figure 20. Close-up view of energy flow around a metal wedge. (a) $\langle S_x \rangle$ (10 times magnified) around a Ag corner. (b) $\langle S_y \rangle$ around a Ag corner. (c), (d), Energy flow vectors: distribution of Poyning vectors around a Ag corner (c) or a PEC corner (d). Note that for both Ag and PEC corner cases the energy flow in the near field region of front surface (x < ~60 nm; y = 0 to 20 nm) is directed left, being funneled around the corner. (e), (f), Poynting vector $\langle S_x \rangle$, and its components $\langle S_x \rangle_{diff}$ and $\langle S_x \rangle_{cross}$ around a PEC corner: scanned along the x-direction at y = 0 (e) or along the y-direction at x = 0 (f). Amplitudes (g) and phases (h) of $E_{y,diff}$, $H_{z,diff}$ and $H_{z,pl}$ fields scanned along the x-direction at y = 0.

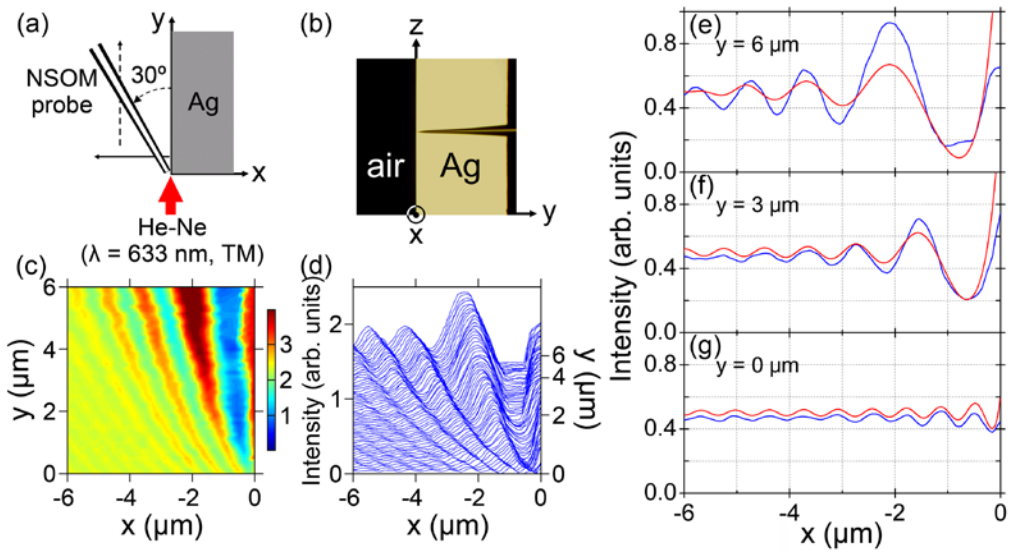


Figure 21. Experimentally measured energy flow distribution around a Ag wedge. (a) Schematic of scanning probe measurement. (b) Optical micrograph of a scanning probe aligned to the edge of a Ag wedge. (c) Color map of measured energy flow distribution (primarily $<S_y>$ component). (d) Measured scan profiles. The coordinate (0,0) corresponds to the corner point. Note the depleted-energy-flow region [dark blue in (c)] along the glancing angle direction near the sidewall (x = 0). (e), (f), (g), Energy flow distributions scanned along the x-direction in the 2nd quadrant: comparison between the experimentally measured data (blue) and a FDTD simulation result calculated at 633 nm wavelength (red). The coordinate (x = 0, y = 0) corresponds to the corner point.

2.2 PHASE RELATIONSHIP OF BOUNDARY DIFFRACTION WAVES

Surface plasmons (SPs) can be excited on a metal surface once the momentum mismatch with the incident free-space wave is properly compensated [1]. A metal wedge structure is one of the fundamental structures that can satisfy the excitation condition of SPs. A metal wedge can diffract the incident wave and generate boundary diffraction waves which play key roles in exciting SPs on metal surfaces [5,6]. Therefore, the study of the boundary diffraction wave for a metal wedge is important. The physical mechanism of a metal wedge responding to an incident optical wave, however, is not essentially studied, even though it can be viewed as a basic element of many plasmonic structures [12,34,35]. In this work we focus on a two-dimensional (2D) metal wedge structure which has a 90º metallic wedge with two semi-infinite surfaces. We develop a fundamental model that can explain the physical mechanism of otpical interactions with the structure and elucidate the essential features of the boundary diffraction waves generated by the structure. With a well established model and the analytical analysis, we provide a method to extract the phase change of the metal reflecting a normal incident optical wave. The FDTD simulation [3] and experimental measurement results confirm the validity of the model. Applying the phase model to a Ag wedge structure, the reflection phase change at the metal/air interface was extracted.

2.2.1 Boundary diffraction wave for a metal wedge and a model for the diffraction

When an optical light is incident to a metal wedge structure normal to one of the surfaces (referred to as the front surface), the light will reflect back at the front surface and diffract away from the corner. Since the diffracted wave is strongly boundary dependent, the wave is termed as

boundary diffraction wave. Generally the boundary diffraction wave is generated due to the discontinuity that the incident wavefronts encounter. For a metal wedge, the boundary diffraction wave can be viewed as consisting of three components generated from different mechanisms. The wave component with planar wavefronts reflected by the front surface only occupies the half-space on the direct incidence/reflection side, and with this categorization the wave component with non-planar wavefronts generated by the reflection process is termed as boundary diffraction wave. This latter wave component corresponds to the first one of the three boundary diffraction waves described above. Besides the reflection and boundary diffraction at the front surface, the light will continue to propagate along the vertical sidewall of the metal wedge. On the transmission side, the incident planar wavefront suddenly becomes of half-infinite due to the blocking at the wedge surface. Due to this edge discontinuity, the propagating planar wave will generate boundary diffraction waves that radiate away from the corner. The non-planar wavefronts that are generated on the transmission side correspond to the second boundary diffraction component. The boundary diffraction being generated on and heading toward the metal side must reflect back to the air side by the vertical surface, and this corresponds to the last component of the three boundary diffraction waves. Figure 22 shows E_y (the electric field projected along the y-direction) of the three components of boundary diffraction waves. The blue line represents the component contributed by the reflection of the front surface, the red line is the one directly generated by the propagating planar wave on the air side, and the green line is for the reflection of the one on the metal side by the vertical surface. Corresponding H_z for the last diffraction component is also shown, from which we can see that the phase difference between E_y and H_z is zero along the x-direction but is non-zero along the y-direction.

As discussed above, the boundary diffraction is generated whenever an incident wave encounters discontinuity. To see this more clearly, we can use a model with a semi-infinite H_z source combined by multiple line sources (point sources when viewed on the x-y domain) extending to one direction infinitely to simulate the current redistribution. Actually, a perfect planar wave can be viewed as an array of infinite H_z sources (extending to both opposite direction infinitely). For a metal wedge, due to the reflected planar part on the metal side and the propagating planar part on the air side, it is equivalent that one semi-infinite H_z source lies on the front surface of the metal and another semi-infinite H_z source lies on the same plane but occupies the air side. We use FDTD method to simulate these two cases separately and thus we can see each mechanism. In the simulation we choose the metal as the perfect electric conductor (PEC) to simplify the modeling process. Since the source lying on the PEC surface will be reflected back and add to the propagating direction causing doubling of the amplitude of the H_z field (no phase change for the reflection of PEC), we reduce the magnitude of the source to half for the initialization. Figure 23(a) shows the simulation H_z field distribution generated by the model for the source lying on the PEC surface and Fig. 23(b), 23(c) and 23(d) show the field distributions of H_z, E_y and H_z after subtracting the planar field part for the source in free space without metal structure. In Fig. 23(a), the 4th quadrant is occupied almost the planar wave part and the 2nd and 3rd quadrants are filled by the quasi-circular boundary diffraction wave part. In Fig. 23(b), in the 1st and the 4th quadrants the boundary diffraction wave is symmetric, and the planar wave part in the 2nd and 3rd quadrants is also symmetric. Since the planar wave part does not have E_y component, the total E_y is purely circular boundary diffraction wave, as shown in Fig. 23(c). If we subtract the planar part in the 2nd and 3rd quadrants in Fig. 23(b), the left part becomes similar to the boundary diffraction wave in the 1st and 4th quadrants, but the phase is anti-symmetric. The

field discontinuity originates from the assumption of a step function profile of planar wave on the left side. Figure 23(d) shows the boundary diffraction wave $H_{z,diff}$. If we place a metal wedge to the 1st quadrant, we can imagine that the wave cannot propagate into the metal, and the boundary diffraction wave shown in Fig. 23(b), 23(c), and 23(d) in the 1st quadrant will be reflected to the other space, mostly in the 2nd quadrant. If we are focusing on the 2nd quadrant including the border between the 2nd and 3rd quadrants ($x < 0$ and $y = 0$), for the case of a metal wedge, the boundary diffraction has totally three components as analyzed above. If the metal is PEC, whose reflection coefficient is 1, implying no phase shift upon reflection, the reflection part from the 1st quadrant will cancel the part in the 2nd quadrant. The total boundary diffraction wave in the 2nd quadrant would then be the same as the result in Fig. 23(a). For non-PEC metal case, the reflection would not cancel the existing part due to the dependence of the reflection on the dielectric constant. Especially, on the vertical surface, when the reflection coefficient is -1 (along the glancing angle direction to the vertical wall), the boundary diffraction wave would be doubled and the total boundary diffraction will cancel the incident planar wave along that direction [36]. This is the origin of the field depletion near the vertical surface. This model elucidates how a magnetic polarized plane wave hitting on a metal wedge generates the boundary diffraction waves.

We also can analytically derive the boundary diffraction wave by using the model. We can start from the unit element of the semi-infinite H_z source. A line source with magnetic polarization can be described by $\tilde{H}_{z,line} = H_{z,line} \frac{e^{ikr}}{\sqrt{kr}}$, where $H_{z,line}$ is the density of the amplitude of the cylindrical wave generated by the source. From FDTD simulation, the relationship of $H_{z,line}$ and $H_{z,pl}$ is determined to be as follows: $H_{z,line} = 4H_{z,pl}$ per micron at 650 nm wavelength. For the semi-infinite H_z source in free space and at $x < 0$ and $y = 0$, as shown in Fig.

23(b), the total H_z field at (x, y) can be written as $H_z(x,y) = \int_{-\infty}^{0} H_{z,line} \frac{e^{ikr_0}}{\sqrt{kr_0}} dx_0$ where $r_0 = \sqrt{(x-x_0)^2 + y^2}$ is the distance from the unit element located at (x_0, 0) to the point at (x, y) and k is the wavevector of the wave in free space. Especially, for $x > 0$ and $y = 0$, $H_z(x,0) = H_{z,diff}(x,0) = H_{z,line} k^{-1/2} \int_x^{\infty} r_0^{-1/2} e^{ikr_0} dr_0$. The integral is the result of the complex complementary error function or the complex Fresnel integral [1,18-19,7-10,37], thus the result of integration gives

$$H_{z,diff}(x,0) = H_{z,line} k^{-1} \sqrt{\pi} e^{\frac{1}{4}i\pi} erfc(\sqrt{-ikx}) = 2H_{z,line} k^{-1} F(\sqrt{-kx}) . \quad (5)$$

For $x < 0$ and $y = 0$, $H_{z,diff}(x,0) = -H_{z,diff}(-x,0)$. Figure 24(a) shows the analytical result of H_z by choosing appropriate time phase well matches the FDTD simulation result of $H_{z,diff}$ along the +x-direction. The y component of the electrical field in the far-field can be written as $E_y(x,y) = \int_{-\infty}^{0} \eta H_{z,line} k^{-1/2} (x-x_0) \left[(x-x_0)^2 + y^2\right]^{-3/4} e^{ik\left[(x-x_0)^2+y^2\right]^{1/2}} dx_0 = \eta H_{z,line} k^{-1/2} \int_{\sqrt{x^2+y^2}=r}^{\infty} r_0^{-1/2} e^{ikr_0} dr_0$, where η is the impedance of the wave in free space and is initialized to unity in the FDTD simulation. Similarly, $E_y(r) = 2\eta H_{z,line} k^{-1} F(\sqrt{-kr})$ for $r >> 0$, which proves the angle independence of E_y (that is, purely circularly symmetric).

The FDTD simulated E_y and $H_{z,diff}$ fields scanned along x-direction and along the y-direction show the relation between E_y and $H_{z,diff}$. Along x-direction, due to E_y is even symmetric but $H_{z,diff}$ is odd symmetric, E_y and $H_{z,diff}$ are almost overlapped for $x > 0$ but have π phase difference for $x < 0$, shown in Fig. 24(a). This can be understood from the fact that the boundary diffraction wave is propagating along the ±x-direction in the source plane ($y = 0$), so $E_{x,diff} = 0$

and $E_y = \pm H_{z,diff}$ for $|x| >> 0$, where "+" for $x > 0$ and "–" for $x < 0$. Along the y-direction, E_y and $H_{z,diff}$ have a phase difference of $\Delta\phi_{diff}$, which is shown in Fig. 22 and Fig. 24(b). This can be analyzed by thinking that the boundary diffraction wave is propagating almost along $\pm x$-direction in the plane touching the source edge ($x = 0$), so $E_{x,diff} = \pm H_{z,diff}$, where "+" for $y < 0$ and "–" for $y > 0$, and E_y and $H_{z,diff}$ have the phase difference of $\Delta\phi_{diff}$. Therefore, since E_y is purely circular (in Fig. 23(c)), $H_{z,diff}$ is just quasi-circular (in Fig. 23(d)), and the phase difference between $H_{z,diff}$ along the x-direction at $y = 0$ and that along the y-direction at $x = 0$ is also $\Delta\phi_{diff}$. From Fig. 24(b), the difference of 1st (or 2nd) peak locations between E_y and H_z is 50°. The value of $\Delta\phi_{diff}$ can be derived analytically as will be discussed below.

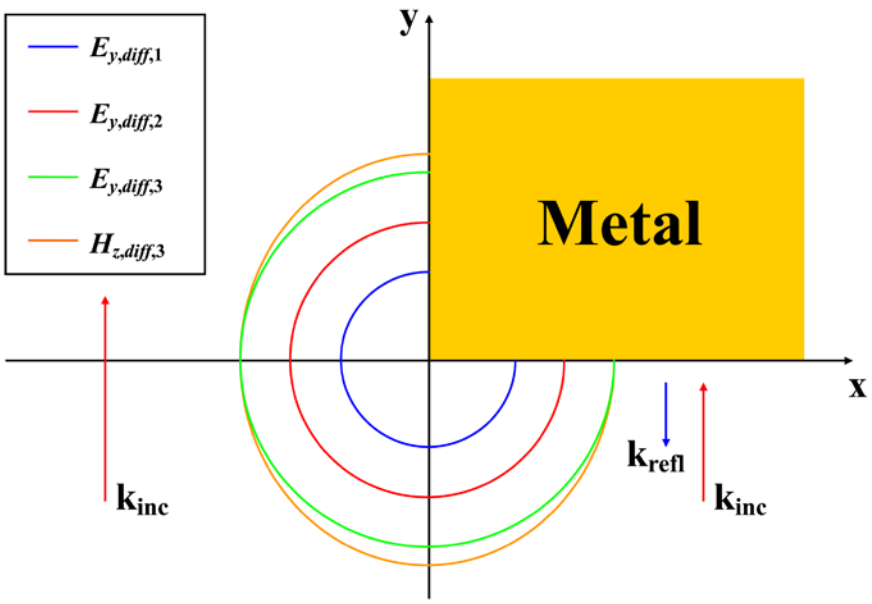


Figure 22. Sketch of all three components of boundary diffraction waves.

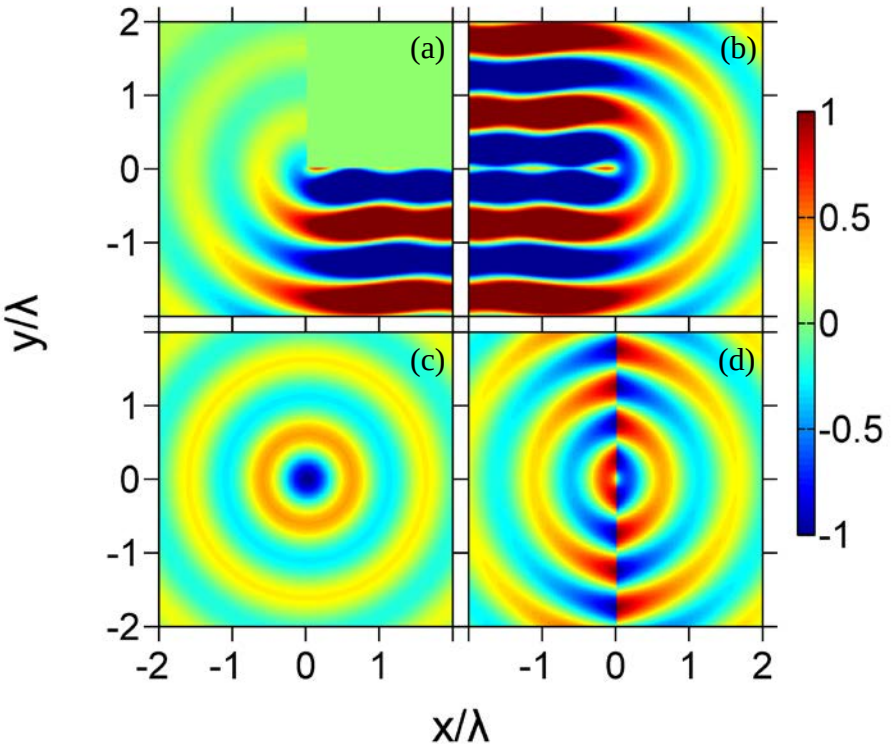


Figure 23. A model for boundary diffraction wave. Snapshot images of field distributions calculated by FDTD. (a) H_z (total) for a half-plane H_z source placed on the bottom face of a perfect electric conductor wedge. H_z (total) (b), E_y (c), and $H_{z,diff} = H_z$ (total) – $H_{z,pl}$ (d) fields for a half-plane H_z source placed in free space.

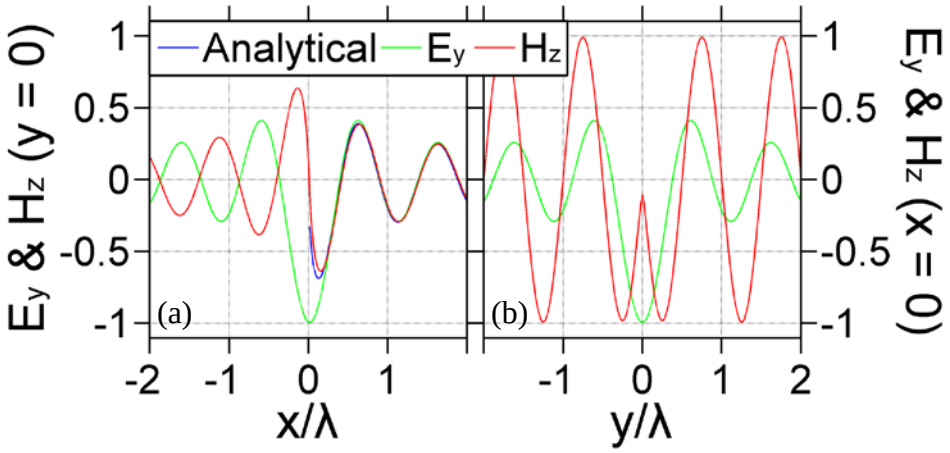


Figure 24. A model for boundary diffraction wave. Horizontal and vertical scans for snapshot images of field distributions calculated by FDTD. (a) Horizontal scan and (b) vertical scan of diffraction fields (analytical $H_{z,diff}$ and simulated E_y and $H_{z,diff}$) for a half-plane H_z source placed in free space.

2.2.2 Phase extraction of normal reflection and measurement

The field profiles along the negative *x*-direction in the plane of front surface ($y = 0$) have been established clearly either by simulation or analytically, and the total field profiles can be measured experimentally. From these results the normal reflection coefficient of the metal can also be determined. A model to extract the reflection phase shift is developed as follows. From our analytical model and the approximation of error function in the intermediate- to far-field: $erfc(\sqrt{-ikr}) \sim \frac{e^{ikr}}{\sqrt{-i\pi kr}}$ ($r >> 0$), for a semi-infinite source in free space (Fig. 23(b), 23(c) and 23(d)), $H_{z,diff}(x,0) \sim \frac{H_{z,line}}{k}\frac{e^{-ikx+i\phi_{effdiff}}}{\sqrt{-kx}}$ ($|x| >> 0$) by scanning along the negative *x*-direction in the front surface plane, where $\phi_{effdiff} = \pi/2$ is the effective phase of the boundary diffraction at $x = 0$ (note that the above approximation is no longer valid for $x \rightarrow 0$, so "effective" is used here) and $H_{z,line}/k$, typically, as mentioned above, can be estimated to be $H_{z,line}/k = 0.4H_{z,pl}$ for the wavelength of 650 nm. At $x < 0$ and $y = 0$, the plane wave source gives $H_{z,pl}e^{i(ky+\phi_{pl})}\Big|_{y=0} = H_{z,pl}e^{i\phi_{pl}}$, where ϕ_{pl} is the phase of the incident plane wave at $y = 0$, therefore the total magnetic field $H_z(x,0) = H_{z,pl}e^{i\phi_{pl}} + 0.4H_{z,pl}\frac{e^{-i(kx-\phi_{effdiff})}}{\sqrt{-kx}}$. From the continuity condition of the propagating planar wavefront and the boundary diffraction wavefront generated in the 1st quadrant (in Fig. 23(b)) along the $+y$-direction at $x = 0$, the phase of the diffraction wave so at the starting point ($y = 0$) is equal to ϕ_{pl}. This phase value can be determined by taking the limit of $H_{z,diff}(0,y)$ for $y \rightarrow 0$. Recall our model, $H_{z,diff}(0,y) = H_z(0,y) = H_{z,line}k^{-1/2}\int_y^\infty r_0^{1/2}(r_0^2 - y^2)^{-1/2}e^{ikr_0}dx_0$ for $x = 0$ and $y > 0$, so

$\phi_{pl} = \angle \lim_{y \to 0} H_{z,diff}(0, y) = \pi / 4$. Therefore, the phase retardation of the diffraction wavefront along the +*x*-direction with respect to that along the +*y*-direction is $\Delta\phi_{diff} = \phi_{effdiff} - \phi_{pl} = \pi / 4$. For a metal wedge filled in the 1st quadrant, as discussed above (in Fig. 22), the boundary diffraction wave (at $x < 0$ and $y = 0$) consists of three components, the phases of which can be analyzed as follows. The first component (in Fig. 22) is associated with the reflection planar wave generated by the front surface of a metal wedge. This reflection has an extra phase retardation of ϕ (the phase of normal reflection coefficient of metal in free space for magnetic fields throughout the work, and the corresponding phase for electric fields should have a π shift). The second one is directly generated by the incident propagating planar wave passing through the wedge, in which the π phase difference exists between any point at the metal side (the 1st quadrant) and its mirror point at the air side (the 2nd quadrant). The last one is the reflection of the second one at the metal side by the vertical surface, where the extra phase retardation of ϕ occurs. Therefore, the phases of three components can be written as $\phi_{pl} + \phi + \Delta\phi_{diff}$, $\phi_{pl} + \pi + \Delta\phi_{diff}$, and $\phi_{pl} + \phi + \Delta\phi_{diff}$. The amplitudes of three components can also be determined as follows. The boundary diffraction wave from the semi-infinite source in free space (Fig. 23(b)) is propagating in both directions whereas in the the metal wedge case the boundary diffraction into metal is blocked. This different configuration results in a factor of 2 difference in the boundary diffraction field amplitude on the plane of y = 0. By considering the reflection loss of the amplitude for the first and last components, the amplitudes of all three components are 0.5ρ, 0.5, and 0.5ρ, where ρ is the amplitude of normal reflection coefficient. Therefore the *y* component of time-averaged Poynting vector is

$$\left\langle S_y(x,0)\right\rangle = -0.5\,\mathrm{Re}[E_x(x,0)\times H_z^*(x,0)]$$
$$= \left\langle S_{y,pl}\right\rangle\left\{1+\frac{0.2}{\sqrt{-kx}}\left[2\rho\cos(kx-\phi-\Delta\phi_{diff})-\cos(kx-\Delta\phi_{diff})\right]\right\}. \quad (6)$$

At the peak locations, noted as x_m, the derivative of $\langle S_y\rangle$ with respect to x vanishes, $\partial\left\langle S_y(x,0)\right\rangle/\partial x\Big|_{x=x_m} = 0$, so

$$\frac{2\rho\cos(kx_m-\phi-\Delta\phi_{diff})-\cos(kx_m-\Delta\phi_{diff})}{2\rho\sin(kx_m-\phi-\Delta\phi_{diff})-\sin(kx_m-\Delta\phi_{diff})}+2kx_m=0. \quad (7)$$

The phase ϕ can be solved once we know ρ and any peak location from simulation or experiment data of the total power at the air side in the plane of the front surface of metal. The measurement of ρ is simple and for most metals, $\rho \sim 1$, so we will use the analytical value in our discussion, which can be calculated by the formula of reflection coefficient: $\rho\exp(i\phi)=\frac{\sqrt{\varepsilon}-1}{\sqrt{\varepsilon}+1}$, where ε is the dielectric constant of metal [6]. From the scan profiles, we can determine the peak locations, and therefore solve the phase retardation of normal reflection by the metal.

To test the validity of the model described above, we use a perfect electric conductor, which can be modeled as $\varepsilon = i\infty$ [6]. Since a PEC is lossless and no phase occurs for magnetic polarized incidence, ρ = 1 and $\phi = 0$. By substituting the analytical value of $\Delta\phi_{diff}$ discussed above into Eq. (6), $\langle S_y\rangle$ calculated by the model (in Fig. 25) has a good match with corresponding FDTD simulation result. On the other hand, we can extract $\Delta\phi_{diff}$ from Eq. (7) by reading the location of the first peak (closest to the corner) from the simulation, which is x_m/λ = -

0.85 (in Fig. 25). So we obtain $\Delta\phi_{diff} = 0.27\pi \sim 49°$. This is very close to the simulation result (50º) shown in Fig. 24(b) and the analytical result (45º).

In our experiment, Ag is used as the metal wedge and the energy flow profiles around the corner are measured by a scanning probe technique. The probe axis is tilted 30 degree off from the vertical direction (Fig. 26(a) and 26(b)), and therefore the measured intensity more closely represents the vertical component of Poynting vector (<S_y>) than the horizontal component (<S_x>). The <S_y> profile (Fig. 26(c)) scanned at $y = 0$ along the negative x-direction shows fringe spacing of 635 nm, which well matches the free-space wavelength of light. The <S_y> profile also shows a perfect match with the simulation result especially for the peak locations.

For the case of a Ag wedge, to calculate the phase ϕ, by taking $\Delta\phi_{diff} = \pi/4$ and the first peak location of <S_y> profile, which is x_m/λ = -0.74 (in Fig 26 or Fig 27(a)) into Eq. (7), we can obtain $\phi = 0.12\pi \sim 22°$. Taking the extracted phase ϕ back to Eq. (6), the calculated <S_y> is achieved and well matches the simulation result in Fig. 27. The analytical calculation of the phase is $\phi = 27°$ (the dielectric constant ε of Ag is used by -17+i1.15 at 650 nm wavelength [17]). The simulation value of $\phi = 22°$ is obtained by direct reading the locations of wavefronts. The phase values from the model, the analytical calculation and the simulation show a good agreement.

When the real part of the dielectric constant is a relatively large negative value, SPs are supported well and will propagate long distance. If the real part becomes close to 0, the propagation length of SPs will decrease dramatically or even no propagation at all [1]. Here we also analyzed the case of poor plasmon-supporting metals, by assuming Re(ε) of metal to -5 or -1.5, the simulated <S_y> is shown in Fig. 27(b). From the simulation, the first peak is located at x_m/λ = -0.66 or -0.55 for Re(ε) = -5 or -1.5, respectively. From the formula calculation of

reflection coefficient, ρ = 0.92 for Re(ε) = -5 and ρ = 0.73 for Re(ε) = -1.5. Similarly, we can solve $\phi = 0.22\pi \sim 40°$ or $\phi = 0.38\pi \sim 68°$, respectively. For comparison, the analytical calculation gives $\phi = 47°$ and the simulation value gives $\phi = 44°$ for Re(ε) = -5 as well as $\phi = 71°$ from the calculation and $\phi = 66°$ from the simulation for Re(ε) = -1.5. All results agree very well.

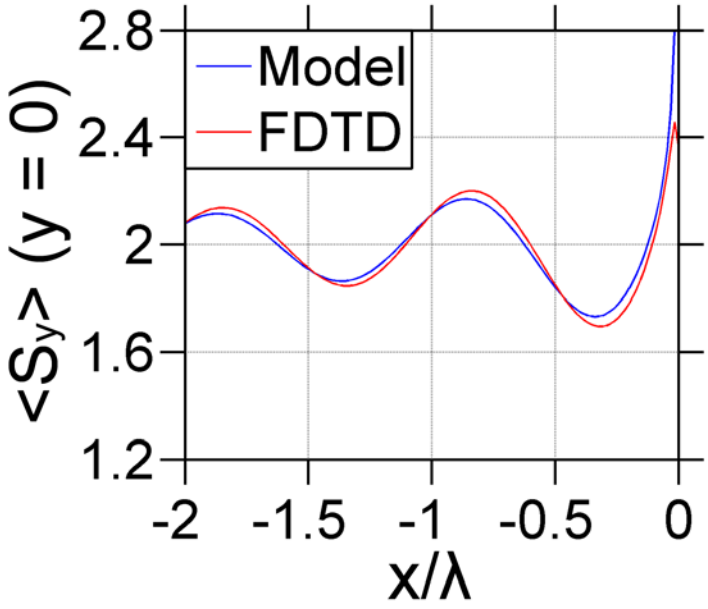


Figure 25. Model and simulation comparison of $\langle S_y \rangle$. $\langle S_y \rangle$ is scanned along the negative x-direction at $y = 0$ for a magnetic polarized plane wave source ($\lambda_0 = 650$ nm) incident to a PEC wedge located in the 1st quadrant.

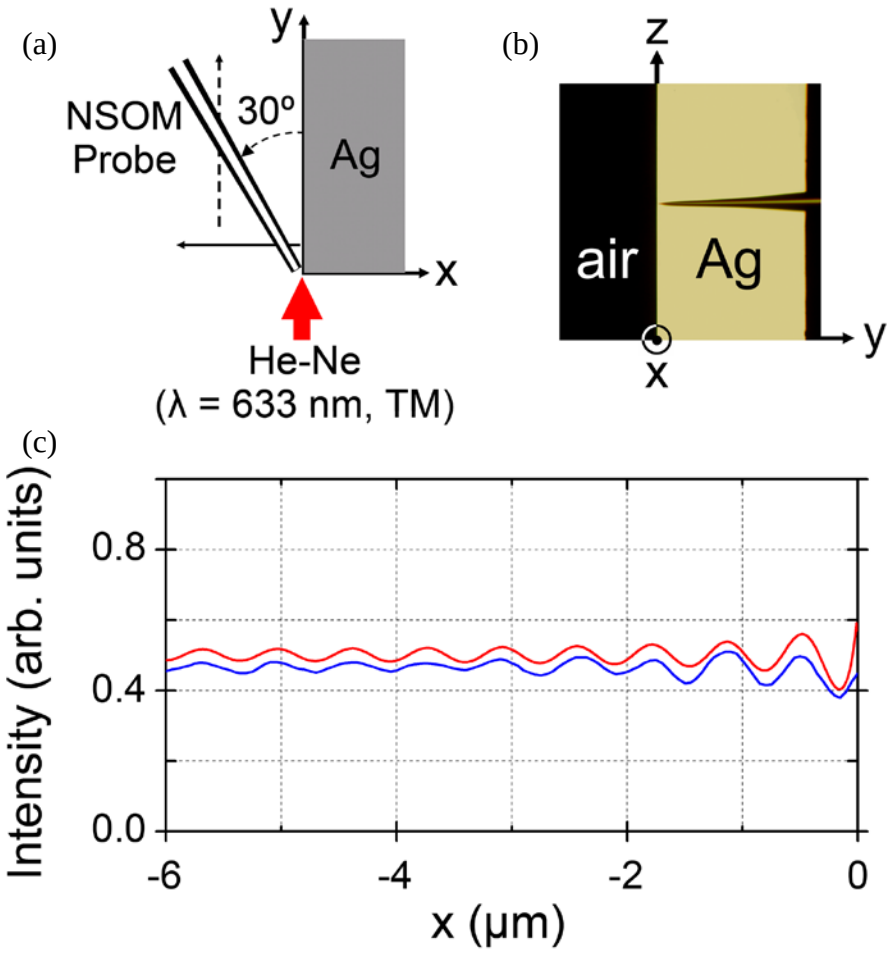


Figure 26. Experimentally measured energy flow profile around a Ag corner. (a) schematic of scanning probe measurement of energy flow distribution around a metal corner. (b) an optical micrograph of a scanning probe aligned to the edge of a Ag corner. (c) Measured intensity profiles of scanning-probe-measured energy flow compare with the simulation result of $\langle S_y \rangle$ at $y = 0$.

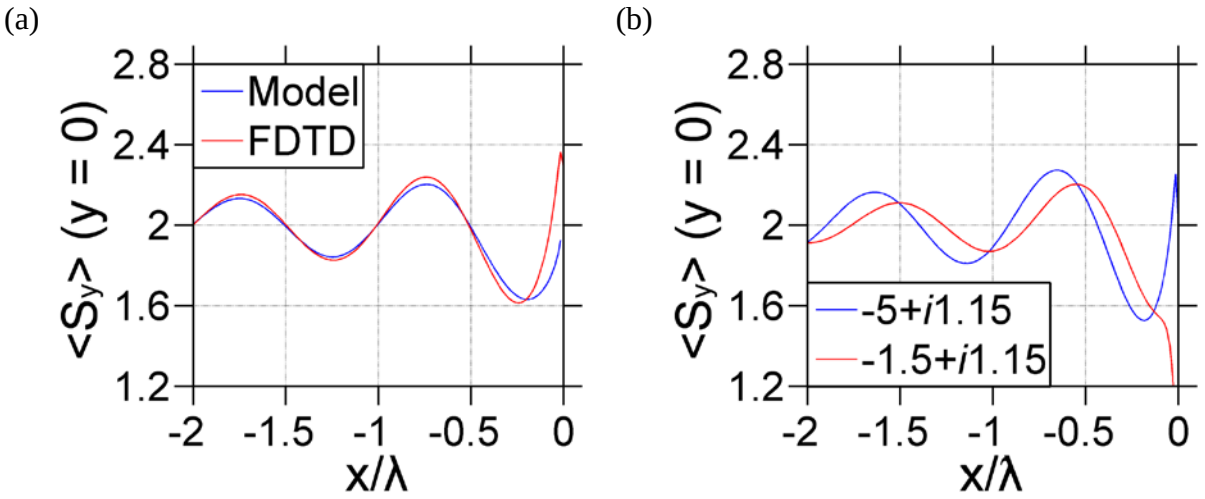


Figure 27. Model and simulation of $\langle S_y \rangle$. $\langle S_y \rangle$ scanned along the negative x-direction at $y = 0$ for a magnetic polarized plane wave source (λ_0 = 650 nm) incident to a real metal wedge located in the 1st quadrant: (a) model and simulation comparison for the case of Ag wedge with ε = -17+i1.15 and (b) simulations for the case of two real metals with ε = -5+i1.15 (blue line) and ε = -1.5+i1.15 (red line) at the wavelength of 650 nm.

2.3 MECHANISM OF SURFACE PLASMON EXCITATION

Proper understanding of the mechanisms of surface plasmon (SP) excitation on a metal wedge structure is essential to further develop related plasmonic structures. As discussed in previous sections, a metal wedge diffracts an incident wave and generates boundary diffraction waves. In this work, we will show how the boundary diffraction wave plays a key role in exciting SPs on metal surfaces [5,6].

For a system without free charge, surface plasmons are characterized by the polarization charges induced at the metal-dielectric interface [27,28]:

$$\rho_P = -\nabla \cdot \mathbf{P} = \nabla \cdot (\varepsilon_0 \mathbf{E}) = \varepsilon_0 \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} \right). \tag{8}$$

For a flat surface without any wedge, the tangential component of electric field is continuous, and the normal component and its derivative determine the polarization charge. For a metal wedge with two perpendicular surfaces, with a planar wave normally incident to one surface, e.g. the front surface, the electric field is therefore normal to the vertical surface. As we analyzed before, the boundary diffraction wave has both the *x*- and *y*- components of electric field that are normal to the vertical and front surfaces, respectively. From this observation we hypothesize that the SPs on the front surface are generated by the boundary diffraction wave and the SP on the vertical surface is generated by the incident and boundary diffraction waves. Next, by employing the model established above and by performing FDTD analyses on some specially designed structures we illustrate the roles played by the boundary diffraction waves in exciting SPs.

Figure 28 shows a FDTD simulation result with a TM-polarized planar wave normally incident to a Ag metal wedge (Fig. 13). The field profiles are scanned along the front surface and vertical surface planes (Fig. 28(e), (f)). In Fig. 28(e), the field profiles are the results obtained by subtracting the planar (incident and reflected) fields from the total fields. The profile shows a superposition of the boundary diffraction wave and the SP wave propagating along the x direction. Since the boundary diffraction along the x direction decays quickly, in the intermediate- to far-fields the SP wave becomes dominant. In Fig. 28(f), as discussed before, field depletion occurs due to the boundary diffraction wave and its reflection by the vertical surface canceling the incident planar wave: even though it is not perfect near the corner, the field profiles are mostly dominated by SPs. From the result we can calculate the efficiency of SP excitation along the vertical surface. The SP excitation on the vertical surface is found to be 4~5 times stronger (more efficient) than that on the front surface. From Eq. (8) and its analysis, SPs are excited on the front surface by the normal component of the electric field Ey. Let's place a half-plane Hz source on the front surface at the air side ($x < 0$ and $y = 0$, in Fig. 29). By choosing a proper phase, the boundary diffraction wave in the 2nd quadrant can be made the same as shown in Fig. 28. Note the boundary diffraction waves for both cases have opposite phases since the waves generated by the half-plane source propagate in the opposite direction in the 3rd quadrant. From Fig. 29(e), the SP excited on the front surface also has an opposite phase compared to Fig. 28(e). It indicates that the SP on the front surface is excited by the boundary diffraction wave component E_y since the incident and reflected waves are planar without E_y component. Because for both cases the E_x components have the same phases along the $+y$-direction in the 2nd quadrant, the SPs excited along the vertical surfaces are also identical, as shown in Fig. 28(f) and Fig. 29(f). Instead of connecting the half-plane source to the corner, we

can separate the source a little bit away from the corner, and the source becomes equivalent to the source shown in Fig. 15(b), (c), and (d). The field profiles along the front surface calculated for different distances between the source and the corner are shown in Fig. 30(a). We can see that the phases of the SPs excited by the *y* component of boundary diffraction waves are identical but the amplitudes attenuate as the distance increases over the wavelength of the incident wave in free space. According to the previous analytical result, the boundary diffraction wave generated by the semi-infinite Hz source in free space is described by the complex complementary error function or Fresnel integral along the source plane ($y = 0$). In intermediate- to far-field, it can be approximately expressed by $\sim 1/\sqrt{r}$. This implies the boundary diffraction field attenuates as the square root of distance from the source edge. In Fig. 30(b), the right-most peak values of H_z shown in Fig. 30(a) (dashed line) prove to attenuate in that square root of distance dependence. Figure 30(c) shows H_z along the *y* direction (vertical surface), from which we can see that only when the source touches the corner, the vertical SPs are excited (the period is 630 ~ 640 nm). For all other sources, the vertical SPs are not excited (the periods are 650 nm, 680 nm, and 720 nm, respectively, due to the boundary diffraction and its reflection near the surface).

Since we showed that the *y*-component of boundary diffraction wave plays an essential role in exciting SPs along the front surface, we can anticipate that the SP should vanish if we block the *y*-component propagating along the front surface. Figure 31 shows the case that a PEC half-plane sheet placed on the plane at $y = 0$ and $x < 0$ and a semi-infinite H_z source is placed on the PEC sheet (Fig. 31(a)). In this structure, the boundary diffraction wave cannot propagate into the 3rd and the 4th quadrants. From Fig. 31(e) we clearly see that the SPs on the front surface are suppressed due to the absence of the E_y component, which is normal to the surface. In Fig. 31(f),

SP along the vertical surface is still excited because Ex in the 2nd quadrant is not affected. Conversely, if we reduce the *x*-component of boundary diffraction in the 1st quadrant, as shown in Fig. 32, by using a Ag sheet with 10-nm thickness placed at $x < 0$ and $y = 0$, the SPs excited along the vertical surface are also reduced to 72% (in Fig. 32(f)), which corresponds to the transmission coefficient of the 10-nm-thick Ag sheet. Because the transmitted plane wave generated boundary diffraction wave transmits again through the sheet, SP excited along the front surface by the y component of boundary diffraction wave in the 4th quadrant is also attenuated by 72% once more, which gives the result of 52% (compare Fig. 32(e) with Fig. 28(e)). Quantitatively, SPs are excited by the normal component of electric field (especially, for the front surface, which is just the y component of boundary diffraction wave).

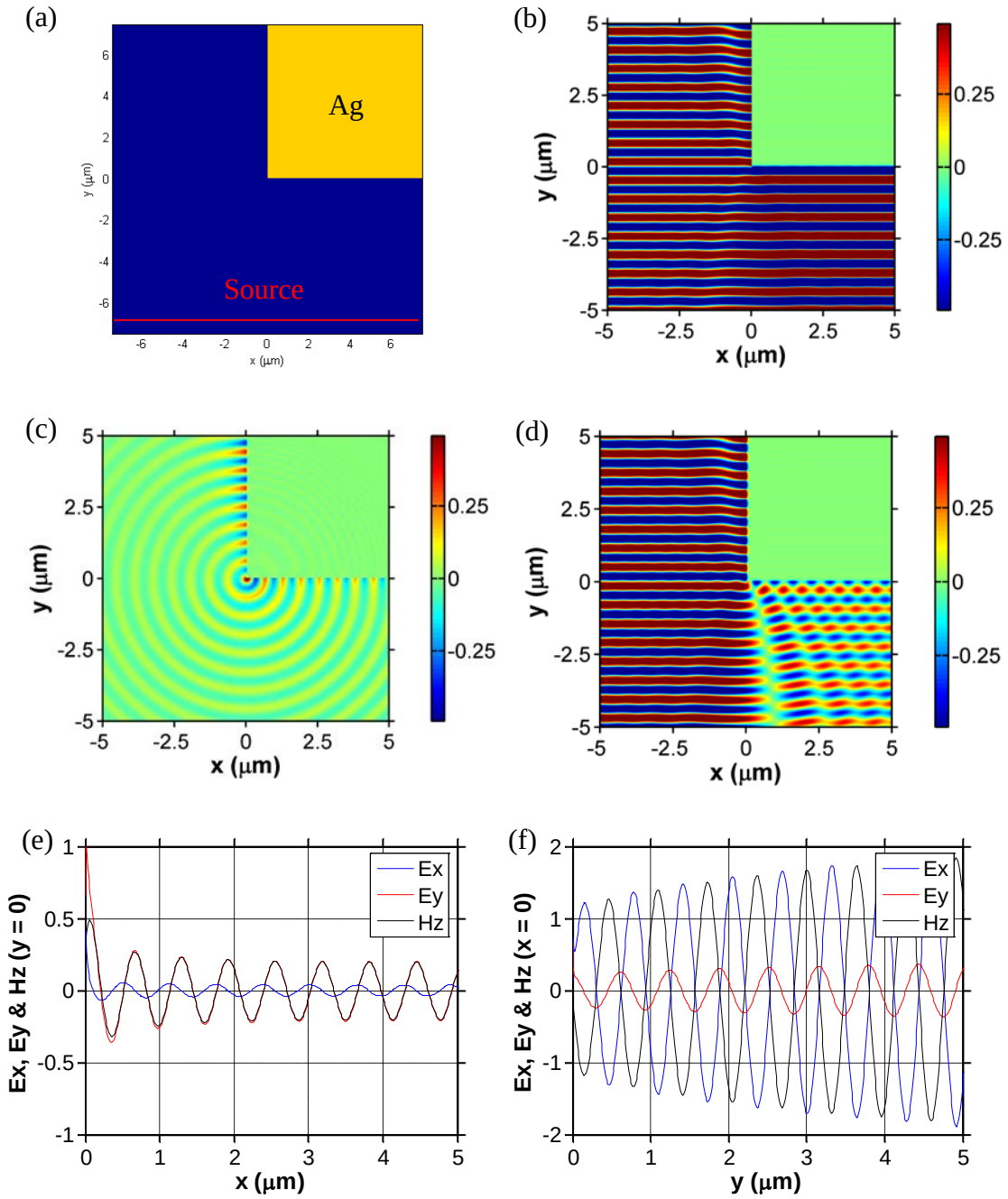


Figure 28. A TM-polarized plane wave source (λ_0 = 650 nm) is striking on a wedge of Ag (ε = -17+1.15*i* at 650 nm). (a) Geometry (b) E_x, (c) E_y, (d) H_z, (e) E_x-$E_{x,pl}$, E_y & H_z-$H_{z,pl}$ scanned along the *x* direction at *y* = 0 and (f) E_x, E_y & H_z scanned along the *y* direction at *x* = 0.

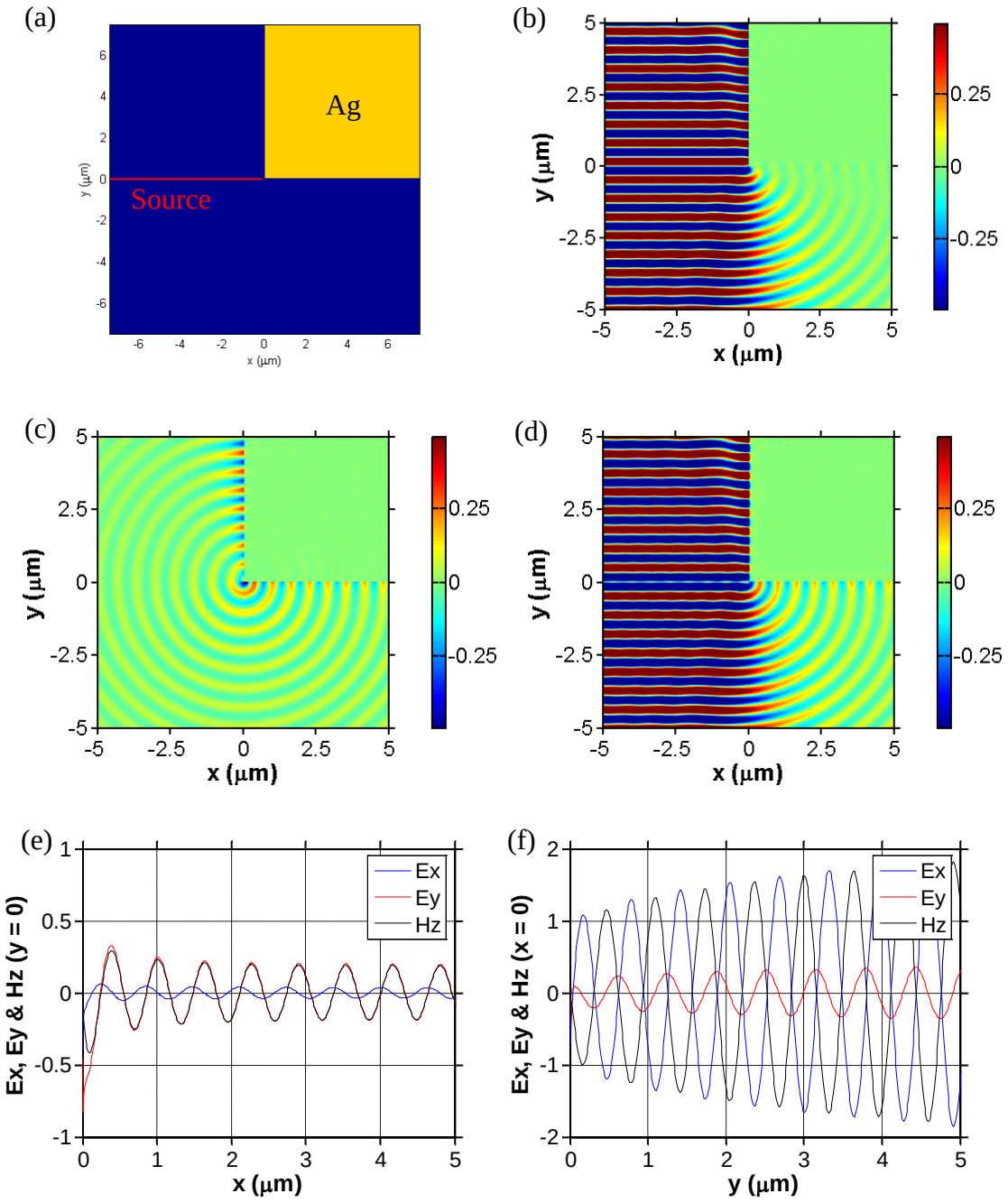


Figure 29. A TM-polarized semi-infinite source (λ_0 = 650 nm) is put on the location with $x < 0$ and $y = 0$ with a wedge of Ag (ε = -17+1.15i at 650 nm) in the 1st quadrant. (a) Geometry (b) E_x, (c) E_y, (d) H_z, (e) E_x-$E_{x,pl}$, E_y & H_z-$H_{z,pl}$ scanned along the x direction at $y = 0$ and (f) E_x, E_y & H_z scanned along the y direction at $x = 0$.

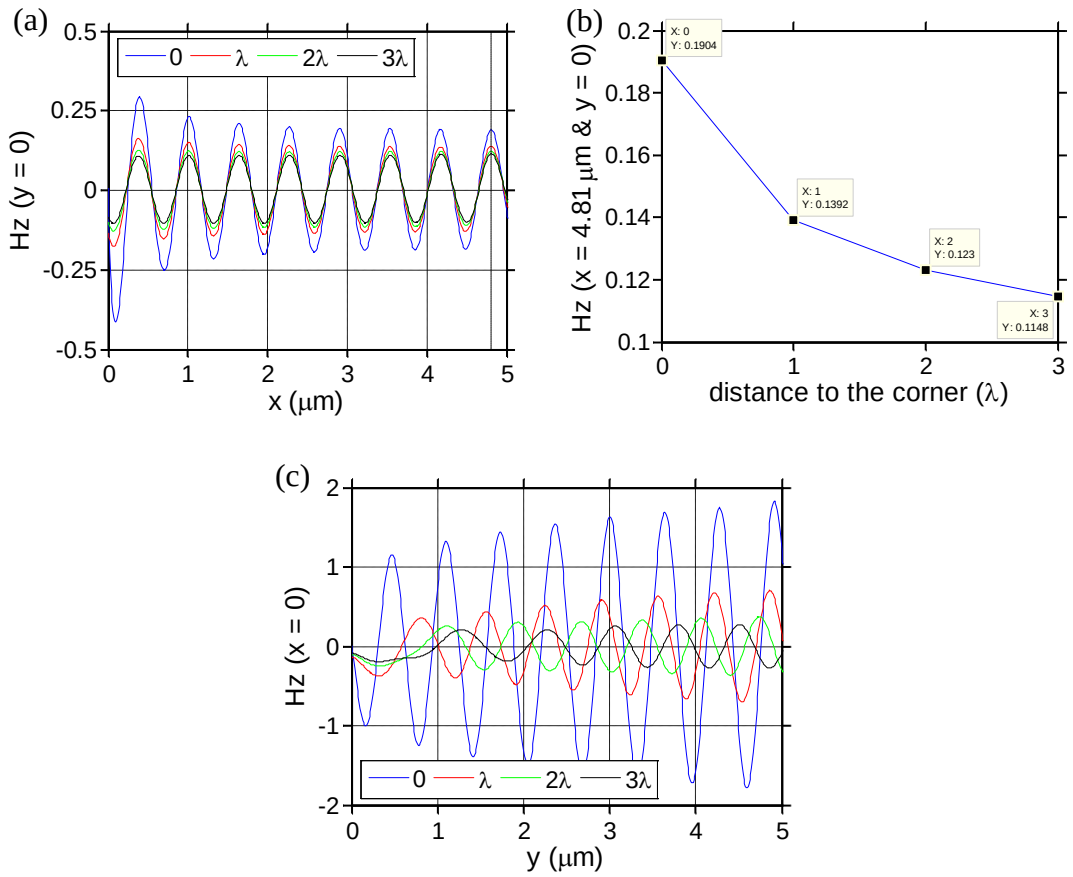


Figure 30. A TM-polarized semi-infinite source (λ_0 = 650 nm) is put on the location with $x < 0$ (shown in Fig. 27), $x < \lambda$, $x < 2\lambda$, or $x < 3\lambda$, and $y = 0$ with a corner of Ag (ε = -17+1.15i at 650 nm) in the 1st quadrant. (a) H_z-$H_{z,pl}$ scanned along the x direction at $y = 0$ and (b) H_z-$H_{z,pl}$ at x = 4.81 μm (dashed line in (a)) and $y = 0$ versus the distance between the semi-infinite source and the corner. (c) H_z scanned along the y direction at $x = 0$.

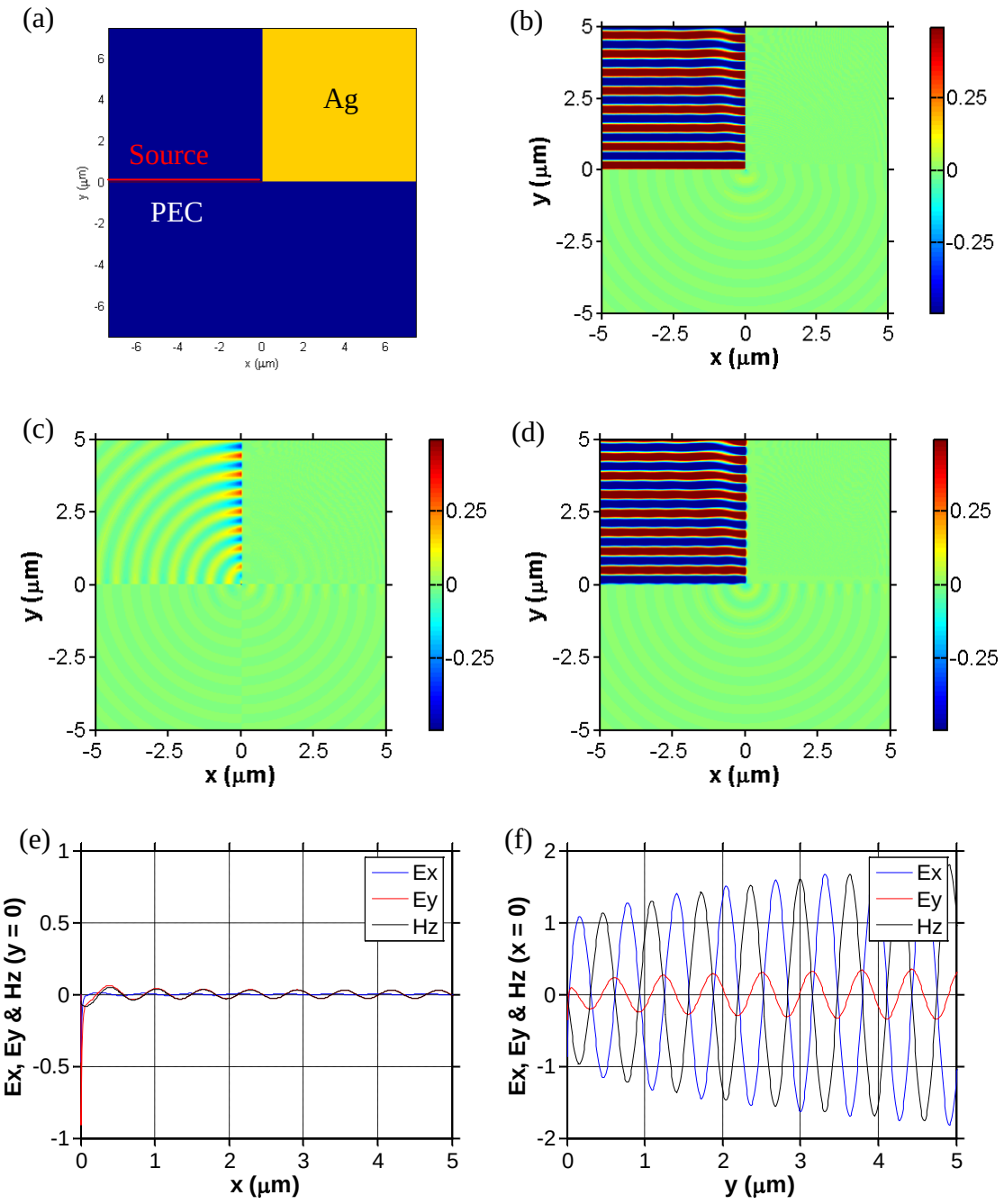


Figure 31. A TM-polarized semi-infinite source (λ_0 = 650 nm) is put on the location with $x < 0$ and $y = 0^+$ with a corner of Ag (ε = -17+1.15i at 650 nm) in the 1st quadrant and a PEC sheet at $x < 0$ and $y = 0^-$. (a) Geometry (b) E_x, (c) E_y, (d) H_z, (e) E_x-$E_{x,pl}$, E_y & H_z-$H_{z,pl}$ scanned along the x direction at $y = 0$ and (f) E_x, E_y & H_z scanned along the y direction at $x = 0$.

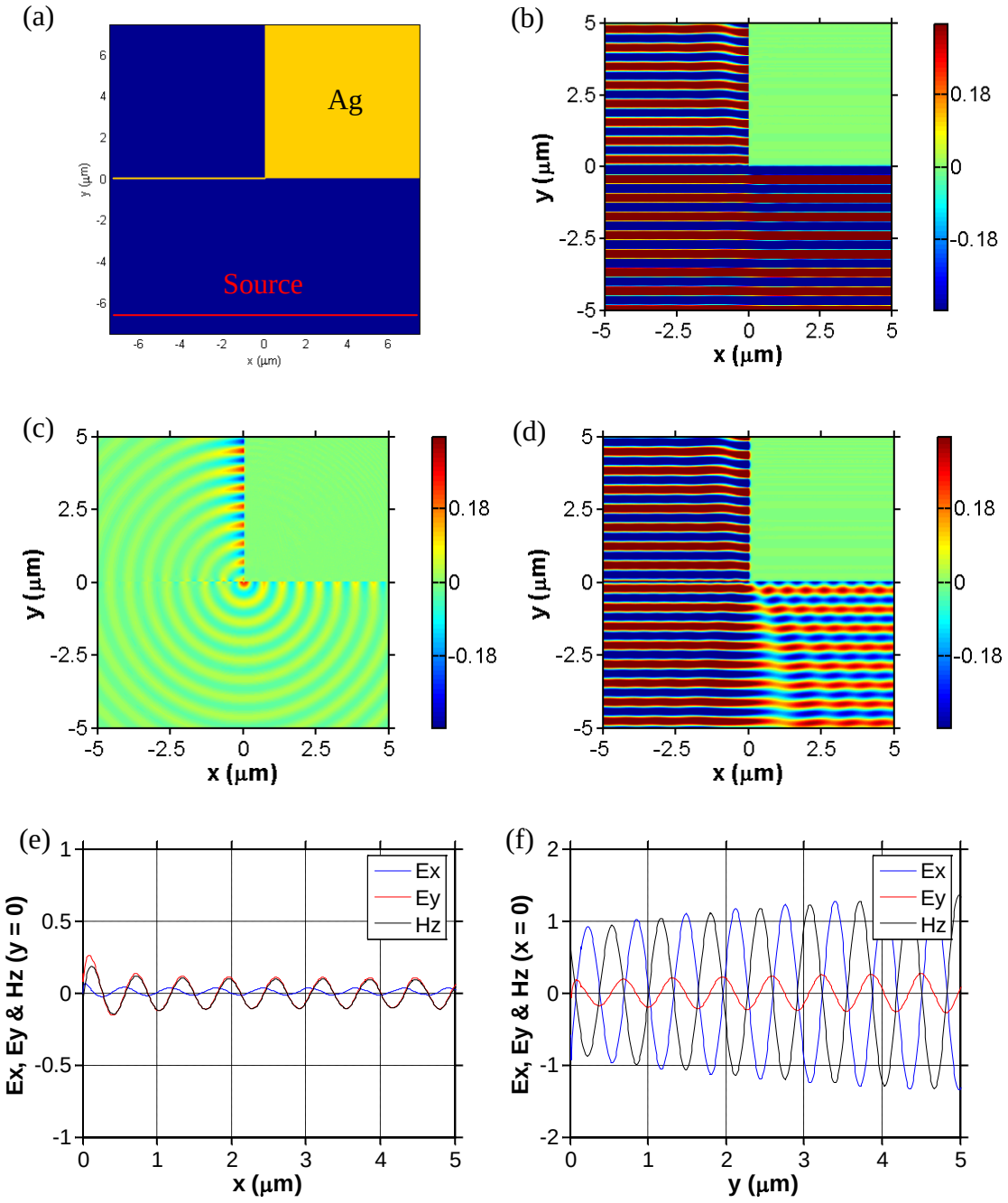


Figure 32. A TM-polarized plane wave source (λ_0 = 650 nm) is striking on a corner of Ag (ε = -17+1.15i at 650 nm) in the 1st quadrant with a Ag sheet (10-nm-thick) located at $x < 0$ and $y = 0$. (a) Geometry (b) E_x, (c) E_y, (d) H_z, (e) E_x-$E_{x,pl}$, E_y & H_z-$H_{z,pl}$ scanned along the x direction at $y = 0$ and (f) E_x, E_y & H_z scanned along the y direction at $x = 0$.

2.4 SUMMARY

The physical nature of electromagnetic interactions of a planar incident wave with a metal wedge is elucidated in this theoretical and experimental study. The boundary diffraction wave is found to play a crucial role in shaping the energy flow distribution around the corner. It is found that a funneling effect is intrinsic to a metal (perfect conductor or real metal) corner structure and is confined to the near field region of a corner. This study confirms that surface plasmons do not play an essential role in funneling at slit edges. A metal wedge's intrinsic capability to redirect the incident energy flow provides a fundamental basis for the commonly observed phenomenon of concentrating an incident energy flow into a narrow channel region. Scanning probe measurement of energy flow distribution confirms presence of a field depletion region along the glancing angle direction of a sidewall. The zero-energy-flow region is caused by destructive interference of boundary diffraction waves near a plasmonic metal surface. The same mechanism provides an explanation of the characteristic radiation pattern (field depletion along the glancing angle direction) of a metal nanoslit structure. A physical understanding of various electromagnetic phenomena associated with a metal wedge structure confirms rich potential of the simple structure as an elemental building block of complex metal nanostructures. The boundary diffraction wave generated by the metal wedge for the planar incident wave toward the front surface of the metal has been studied. A physical model has been developed to analyze the essence of the boundary diffraction. With this model, the phase relationship among the diffraction wave components has been analyzed and the reflection phase retardation of the metal to the normal incidence has been extracted from the FDTD simulation results or experimental data. The analysis and measurement techniques presented in this paper provide a simple and

useful method to measure the dielectric constant or refractive index of a real metal. It may also be expanded to other phase related optical measurement of different materials.

3.0 OPTICAL WAVEFRONTS EMANATING FROM A METAL NANOSLIT

Interaction of light with a nanoapertured metal film has been the subject of extensive study because it produces many interesting phenomena, such as “enhanced” transmission of light through a nanohole array or shaping the spatial or spectral profiles of the transmitted light [11-13,21,22,35,38,39,40,41,42]. The richness of the phenomena stems from the complexity of the way that light interacts with the nanostructures formed in the metal film. Surface plasmons (SPs), collective oscillation of electrons carrying the electromagnetic energy in the form of light trapped at a meta/dielectric interface, can effectively mediate the interactions between metal nanostructures [1]. Unlike the dielectric case, a metal nanostructure can also efficiently interact with free-space radiation, diffracting an incident light and/or coupling the light into surface plasmons (vice versa, decoupling surface plasmons into free-space radiation). This implies the multiplicity of the interaction pathways available on nanostructured metal surface. When properly designed, the effects of near-field interactions between nanostructures can also reach the far-field region through diffraction and constructive interference among them. The spatial and/or spectral profiles of far-field optical transmission through a metal nanoslit array, for example, are known to be governed by various resonances occurring on different sections of metal surface [43]. The roles played by surface plasmons and free-space radiation in the interaction of metal nanostructures, however, have not been clearly understood and have been a subject of debate. The surface-bound wave and free-space radiation behave differently in terms

of propagation constant (wavelength and attenuation), field distribution, etc., and analyzing individual roles and their interplay in an arbitrary structure is considered a challenge.

A single nanoaperture formed in a metal film is a simple and yet the most fundamental structure that can be viewed as a basic building block of aperture-based nano-plasmonic structures. Evolution of optical wavefronts emanating from a metal nanoaperture is of essential interest in studying the plasmonic structures, and yet a detailed understanding is not fully established on how the different wave components (free-space diffraction and surface bound waves) interplay and evolve over the near- to far-field regime [12,13,21,22,35,39,40,41]. In this chapter we report near- to far-field imaging of optical wavefronts emanating from a single nanoslit formed on a thin Ag film.

3.1 PHASE ANALYSIS OF NEAR- TO FAR-FIELD OPTICAL INTERFERENCES

A 80-nm-wide slit was formed with focused-ion-beam etching in a 50-nm-thick Ag film deposited on a fused-silica substrate (Fig. 33(a), left). A TM-polarized laser light (633 nm wavelength) was incident to the substrate side, and the transmitted light was imaged by scanning a nanoprobe (Veeco Aurora NSOM probe 1720-00: 100-nm-thick Al coated; 80-nm aperture diameter) in the horizontal direction at the exit side (Fig. 33(a), right). Figure 33(b) shows some of the scan profiles obtained in the near- to far-field regime using a metal nanoapertured probe. Fringes form and evolve over the entire regime, since the light partially (and directly) transmitted through the thin Ag film interferes with the waves transmitted through the nanoslit [44,45,46]. The fringe spacing increases for larger probe-to-surface distance and/or towards the central region on a given scan. Figure 33(c) shows a two-dimensional (2D) map of the scan profiles over

the entire regime (near- to far-fields): 71 scan profiles are displayed with the fringe amplitude color-coded.

In order to elucidate the interference nature of fringe formation, a schematic of two propagating waves is shown in Fig. 34(a), one emanating from a nanoslit with cylindrical wavefronts and the other directly transmitted through a thin metal film for a planar wave incident from the bottom side. The fringes resulting from constructive interference of the two waves are marked on the cross points of the wavefronts. In this diagram, the (*m*,*n*)-th cross point corresponds to the interference of the *m*-th cylindrical wavefront (*m* = 1, 2, 3,…) radiating from the nanoslit and the *n*-th planar wavefront (*n* = 0, 1, 2, 3,…) directly transmitted through a metal film. After a simple analysis, the coordinates of the (*m*,*n*)-th cross points, ($x_{m,n}$, $y_{m,n}$) can be determined as follows.

$$x_{m,n} = \left[\left(m + \frac{\phi}{2\pi} \right)^2 - n^2 \right]^{1/2} \lambda ; \qquad y_{m,n} = n\lambda \tag{9}$$

Here λ is the free-space wavelength of the transmitted light. ϕ is the phase difference between the directly-transmitted planar wave (ϕ_1) and the nanoslit-transmitted cylindrical wave (ϕ_2), that is, $\phi = \phi_1 - \phi_2$, and $|\phi| < \pi$. In the region far from the slit and yet with relatively small probe-surface distance (*m* >> *n*), the fringe location can be expressed as $x_{m,n} \cong (m + \frac{\phi}{2\pi})\lambda$; $y_{m,n} = n\lambda$. This tells the fringe spacing asymptotically approaches the free-space wavelength, and the fringe locations are off-shifted from the integer multiple of wavelength positions by the amount proportional to phase retardation ϕ.

Interference of two waves of different symmetry (cylindrical versus planar) results in a characteristic fringe pattern (Fig. 33(c) and Fig. 34). The first track on either side from the center corresponds to the cross-points of the *n*-th planar wavefront and the (*n*+1)-th radial wavefront, i.e., the case of $m = n + 1$ in Eq. (9). In general, the *l*-th track comprises the cross-points, ($x_{m,n}$, $y_{m,n}$) with $m = n + l$. On a given track and in the region far from the metal surface ($n >> l$), the fringe location traces approximately a parabolic profile, i.e., $x_l(n) \cong \lambda\sqrt{2nl}$. Figure 34(c) shows a Poynting vector (the *y*-component) distribution calculated from the finite-difference-time-domain (FDTD) analysis of wave transmission through a nanoslit [47]. [The vertical (*y*-) component of the Poynting vector is the main contributor to the scanning probe output, since the probe is aligned normal to the sample surface [39].] Overall the fringe pattern obtained from a FDTD simulation shows a good agreement with the measurement result (Fig. 33(c) and Fig. 34(c)). Figure 34(b) shows a detailed comparison between the cylindrical-and-planar-wave-interference model (blue dashed curves) and the FDTD simulation (red dashed curves) results. While both agree well in the far-field regime, a clear difference is observed in the near- to intermediate-field regime: the FDTD simulation predicts a fringe track shifted inward from that of the two-wave-model. This discrepancy is ascribed to stronger presence of surface plasmon waves near the metal surface compared with the slit-transmitted cylindrical wave (Fig. 34(c)). The fringe spacing near the metal surface is significantly smaller than the free space wavelength as can be seen in the mismatch of the fringe patterns in the near and far-fields. The fuzzy area corresponds to the transition region between the two distinct regimes where either slit radiation or surface plasmon field is dominant over the other. In this region, phase singularity exists at the points where the two fringes completely mismatch [21,45] (see the inset of Fig. 34(c)). The phase singularity indicates that the optical fields of slit-diffraction and surface plasmons cancel

each other and the phase becomes undefined in the local area. (It should be mentioned that the surface plasmon and free-space radiation waves have different wavelengths, therefore no absolute phase value can be defined for both waves in this region.)

For detailed analysis, the three wave components that are expected to have significant presence in the near- to far-fields are expressed as follows. 1) a partial, directly-transmitted wave (TM polarized) through a thin metal layer: $\mathbf{H}_1 = \hat{\mathbf{z}} H_1 e^{iky}$ and $\mathbf{E}_1 = -\hat{\mathbf{x}} \eta H_1 e^{iky}$, where $\eta = \sqrt{\mu / \varepsilon}$. 2) a free-space radiation wave emanating from the slit: $\mathbf{H}_2 = \hat{\mathbf{z}} H_2 e^{i\mathbf{k}\cdot\mathbf{r}}$ and $\mathbf{E}_2 = (-1/i\omega\varepsilon)\nabla \times \mathbf{H}_2$. In the regime where a slowly varying condition (i.e., $|H_2| >> \left|\frac{\partial H_2}{\partial x}\right|, \left|\frac{\partial H_2}{\partial y}\right|$) is satisfied, the electric field vector can be approximated as follows. $\mathbf{E}_2 = (-\hat{\mathbf{x}} \sin\theta + \hat{\mathbf{y}} \cos\theta) \eta H_2 e^{ik(x\cos\theta + y\sin\theta)}$. 3) surface plasmons generated at the slit edges and propagating away from the slit: $\mathbf{H}_3 = \hat{\mathbf{z}} H_3 e^{-\gamma_{sp} y} e^{\pm i k_{sp} x}$ and $\mathbf{E}_3 = (-1/i\omega\varepsilon)\nabla \times \mathbf{H}_3 = \hat{\mathbf{x}}(\gamma_{sp} / i\omega\varepsilon) H_3 e^{-\gamma_{sp} y} e^{\pm i k_{sp} x} + \hat{\mathbf{y}}(\pm k_{sp} / \omega\varepsilon) H_3 e^{-\gamma_{sp} y} e^{\pm i k_{sp} x}$. Here the ± sign corresponds to the surface plasmons propagating along the positive (+) or negative *x*-direction (-), respectively. Regarding the phase relationship of slit-transmitted waves, it should be noted that the free-space radiation and SP waves (H_2 and H_3) are in phase at the slit exit, since they originate from the same surface plasmon wave transmitted through a nanoslit. The region proximal to the slit and the intermediate-to-far-field regime (where the SP fields are negligible) can be well described in terms of interplay of the two wave components, direct transmission through a film $(\mathbf{E}_1, \mathbf{H}_1)$ and nanoslit radiation $(\mathbf{E}_2, \mathbf{H}_2)$. The time-averaged energy flow can then be expressed with the Poynting vector as follows. $\langle \mathbf{S} \rangle = (\mathbf{E}_1 + \mathbf{E}_2) \times (\mathbf{H}_1 + \mathbf{H}_2)^*$. As the probe scans away from the slit ($\theta \to 0$) at a constant probe-surface distance ($y = y_0$), the *y*-component

of the Poynting vector (the main contributor to the probe output) asymptotically approaches the following expression:

$$\left\langle S_y \right\rangle = \eta |H_1|^2 + \eta |H_1||H_2| \cos(kx - ky_0 - \phi) + i\eta |H_1||H_2| \sin(kx - ky_0 - \phi) \tag{10}$$

Here ϕ is the phase difference between the directly-transmitted and slit-transmitted waves as defined above with Eq. (9). The real part of the Poynting vector component, $\mathrm{Re}\left(\left\langle S_y \right\rangle\right)$ corresponds to the scan profile measured with a nanoapertured probe, and the fringe peaks occur at $x = \lambda(m + \frac{y_0}{\lambda} + \frac{\phi}{2\pi})$. Note that at $y_0 = n\lambda$ this formula reduces to the one derived from the two-wave-interference diagram shown in Fig. 34(b). The phase relationship (ϕ) of the slit transmitted wave and the direct transmission determines the exact location of fringes (the offset from the positions at integral multiple of free-space wavelength).

In the vicinity of the slit, the radial wave (H_2) shows faster damping ($1/r^{1/2}$) compared with the SP's exponential decay along the *x*-direction. In the region distant from the slit but near the metal surface, the wave interaction can be described in terms of SP and direct transmission, and the probe output at $y = 0$ can be expressed as follows.

$$\begin{aligned} \mathrm{Re}\left(\left\langle S_y \right\rangle\right) &= \eta |H_1|^2 + \eta |H_1||H_3| \cos(k_{sp} x \mp \phi) \mp \frac{\gamma_{sp}}{\omega\varepsilon} |H_1||H_3| \sin(k_{sp} x \mp \phi) \\ &= \eta |H_1|^2 + \eta |H_1||H_3| A \cos(k_{sp} x \mp \phi \pm \alpha) \end{aligned} \tag{11}$$

Here $\alpha = \tan^{-1}(\gamma_{sp} / k)$. The fringe peaks occur at $k_{sp} x = 2\pi m \pm \phi \mp \alpha$ with the spacing equal to the plasmon wavelength λ_{sp}. Compared with the far-field case, the fringe location is affected by an extra term α, which is separate from the shift caused by the phase retardation ϕ. This extra phase shift originates from the horizontal component of the SP electric field (E_{3x}), which is out-of-phase (by $\pi/2$) with its normal component (E_{3y}) [1,27].

Figure 35 shows a detailed comparison of the scan profiles measured in the near- to intermediate- to far-field regimes (y = 200 nm, 630 nm, and 7.6 μm, respectively) with the corresponding FDTD simulation results. Overall a good agreement is observed in all three regimes. According to the FDTD simulation (Fig. 34(c)), the region outside ~1 μm distance from the metal surface is free from the effect of surface plasmon presence. The fringe locations in this regime are governed by the interference of direct transmission and slit radiation, and can be described by Eq. (9). From the measured peak locations and referring to Eq. (9) with λ = 633 nm, n = 12, and m = 13 to 19, the phase retardation ϕ is estimated to be -86 ± 10 degrees. This negative retardation depicts a relationship that the phase of the nanoslit-transmitted wave leads the directly transmitted wave by ~86 degrees (or equivalently, in terms of wavefronts, trails by ~86 degrees). In the case of near-field scan, the measured fringe spacing is 605 ± 10 nm, showing a reasonable agreement with the surface plasmon wavelength λ_{sp} (613 nm) as expected from Eq. (11).

Next we elucidate the nature of phase evolution of each wave component during transmission through a nanoslit or a thin metal film. For a given planar wave incident from the substrate side, the three waves take different paths, accumulating different amount of phase when measured at the exit surface of nanoslit.

For the case of direct transmission through a metal film (of thickness *d*), the total phase retardation can be calculated from the transmission coefficient of the metal layer: $t = \frac{t_{12}t_{23}e^{ik_2 d}}{1 - r_{21}r_{23}e^{i2k_2 d}}$. Here t_{ij} is the transmission at the interface of the *i*-th and the *j*-th layers, and can be expressed as $t_{ij} = 2\sqrt{\varepsilon_i}/(\sqrt{\varepsilon_i} + \sqrt{\varepsilon_j})$. *i* = 1, 2, and 3 represents substrate (quartz), metal (Ag) and air side, respectively, and ε_i is the corresponding dielectric constant. Similarly, r_{ij} is the reflection coefficient at the *i*-*j* layer interface with the *i*-th layer as the incidence side, and is expressed as $r_{ij} = (\sqrt{\varepsilon_j} - \sqrt{\varepsilon_i})/(\sqrt{\varepsilon_i} + \sqrt{\varepsilon_j})$. The total amount of phase change of direct transmission is calculated to be -50 degrees. [For the metal thickness (50 nm) studied in this work, the single pass transmission in the metal layer is 0.14, therefore the multiple internal reflection effect is insignificant. The main contributions are from the two interfacial transmissions, i.e., from the phases of t_{12} and t_{23}]. The phase change of a slit-transmitted wave comes from the phase accumulation during SP propagation through a slit, Re(k_{sp})*d*. For the case of a 80-nm-wide and 50-nm-deep slit, the phase retardation is estimated to be +35 degrees. Combining the two phase change components, overall the nanoslit transmitted wave leads the directly transmitted wave by 85 degrees of phase, that is ϕ = -85 degrees. Figure 36 shows a FDTD simulation of phase relationship of the two wave components. The wave transmitted through a nanoslit leads the direct transmission by 90 degrees of phase as can be seen from the comparison of wavefront locations at the exit side for the same incident wave. The analytical and simulation results clearly confirm the phase retardation (ϕ = -86 ± 10 degrees) extracted from the measurement data discussed above.

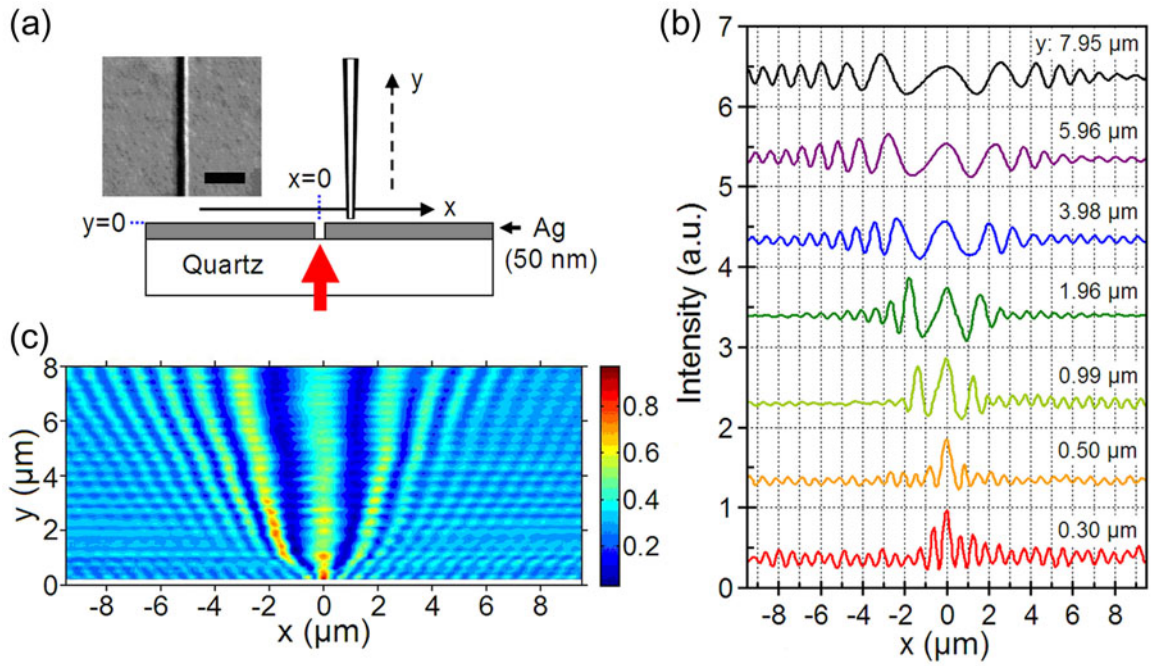


Figure 33. Near- to far-field imaging of optical wavefronts emanating from a nanoslit formed in a thin Ag film. (a) TM polarized laser light is incident to the nanoslit (80-nm wide and 50-nm thick: inset, SEM image; scale bar, 500 nm) from the substrate side, and a nanoapertured scanning probe is scanned along the horizontal direction with a step size of 50 nm. (b) Scan profiles of the interference pattern of slit-transmitted and direct film-transmitted waves. The base line of each scan is shown as a dotted line. (c) 2D map of the interference pattern with the intensity color-coded. Red corresponds to the peak of a fringe and blue represents the valley. The periodic modulation of intensity along the vertical (*y*) direction is ascribed to the Fabry-Perot resonance effect of a local cavity structure formed by the probe tip and the sample surface.

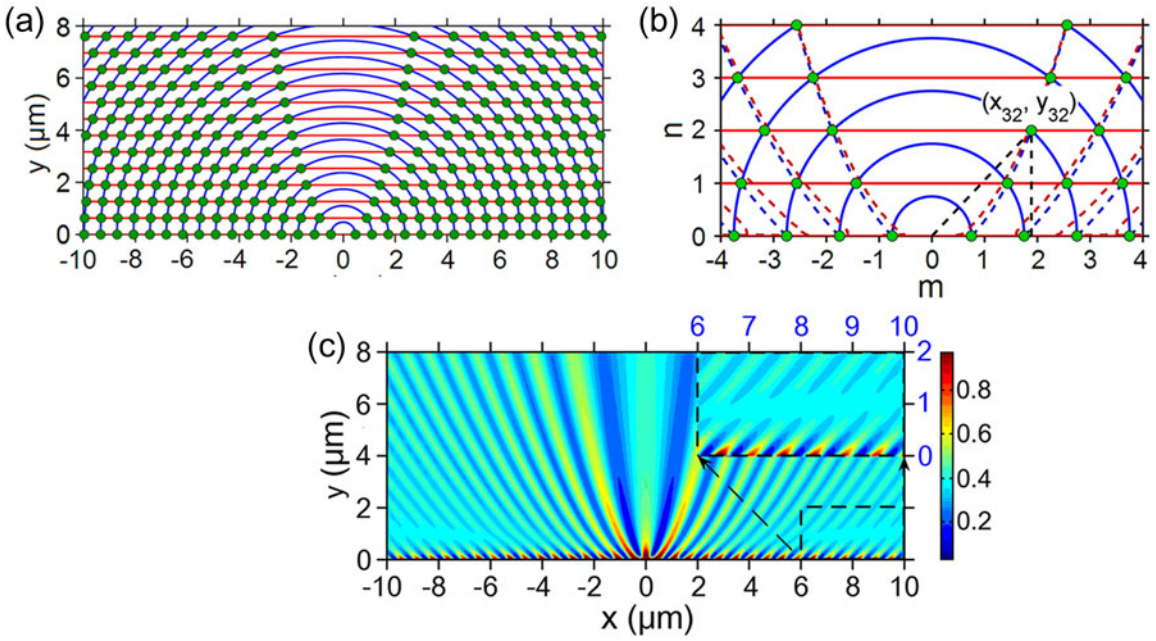


Figure 34. Interference of slit-transmitted and direct, film-transmitted waves. (a) Wavefronts of slit-transmission (cylindrical curves) and direct film-transmission (horizontal lines). (b) A close-up view of the low-order fringe tracks in the two-wave interference model. The blue dashed curves represent the fringe tracks calculated from the two-wave model, and the red dashed curves are a FDTD simulation result. (c) The interference fringe pattern calculated from a FDTD simulation of optical transmission through a Ag nanoslit (80-nm wide and 50-nm thick): The vertical component of the Poynting vector. The inset in the top right part is a magnified view of the bottom right corner region (*x*: 6-10 μm; *y*: 0-2 μm). The dielectric constants of silica and Ag used in this simulation are from References 17.

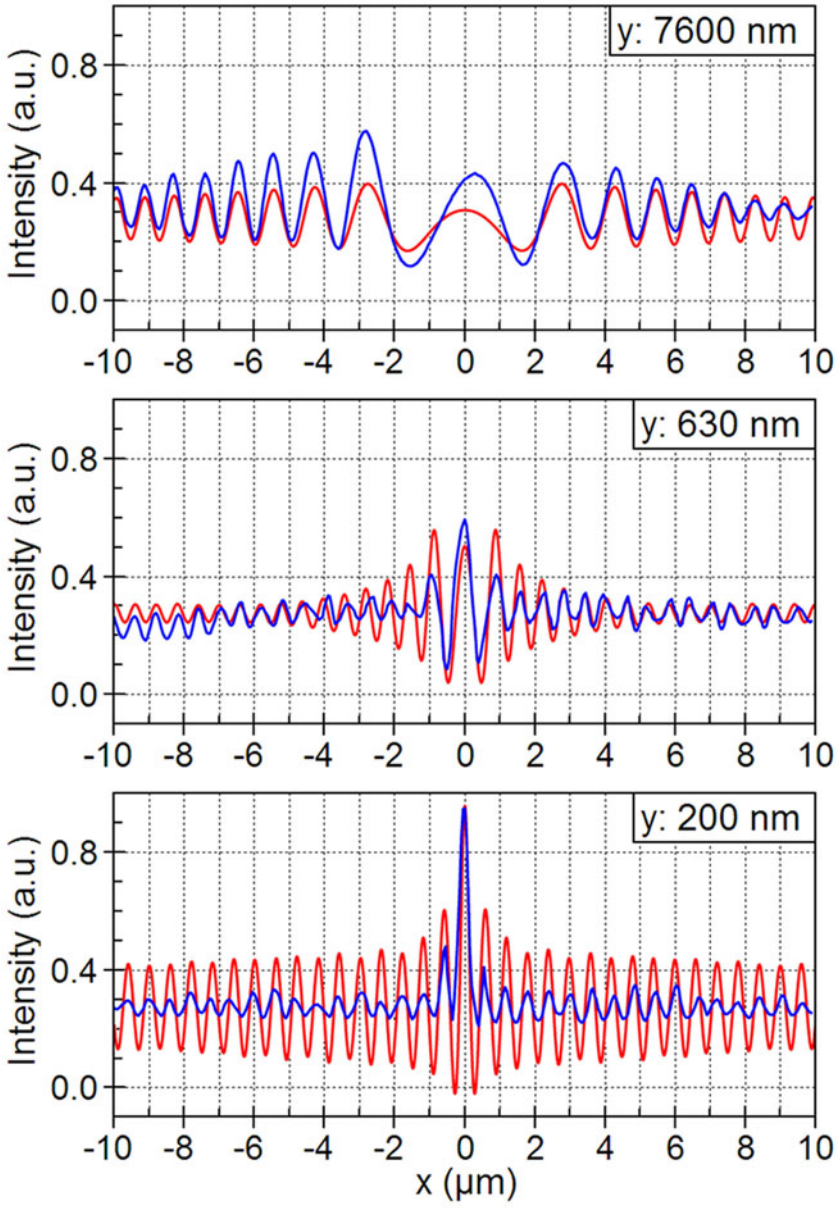


Figure 35. Comparison between the measured scan profiles and FDTD simulation result. The scan profiles were measured with the same slit as in Fig. 33 at probe-surface distance of 7.6 μm, 630 nm, or 200 nm (top to bottom) with a scan step size of 10 nm. Blue curves correspond to the measurement, and red to the FDTD.

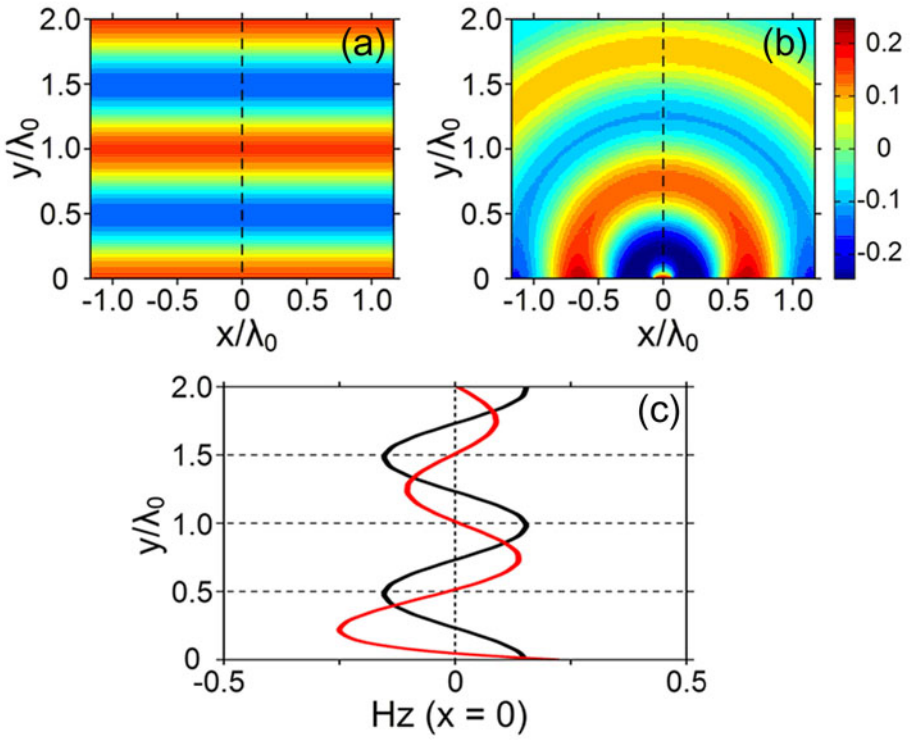


Figure 36. FDTD simulation of phase relationship of slit-transmission and direct film-transmission. The wavefronts of (a) direct film-transmission and (b) slit transmission. (c) Comparison of optical phases of direct transmitted wave (black) and slit transmitted wave (red). The direct transmission was calculated for a 50-nm-thick Ag layer without a slit. The slit radiation was calculated from a nanoslit simulation result by subtracting the direct transmission component.

3.2 AMPLITUDE ANALYSIS OF NEAR- TO FAR-FIELD OPTICAL INTERFERENCES

As discussed above, a single slit on a Ag layer can excite surface plasmons (SPs) at entrance edges and the excited plasmons propagate through the slit. At the exit side, the slit-transmitted SPs are partially decoupled into free space and partially continue to propagate along the horizontal Ag surfaces. When the thickness of Ag layer is small enough, the incident wave can also directly transmit through the metal film. The three wave components can interfere with each other. The phases of the slit-transmitted waves can be retarded compared to that of the direct film-transmitted wave. This is because the SPs propagating inside the slit have different propagation constant than the direct film-transmission case. The phase relationship between them has already been studied as discussed previously. Amplitude is another important information of an electromagnetic wave. Amplitude and phase together can determine a specific wave completely. So the amplitude relationship between slit-transmission and direct film-transmission is also meaningful to study. From both amplitude and phase information of transmission, the properties of thin metal films can be further studied.

To extract the amplitude information of transmitted waves, both the direct film-transmitted waves and slit-transmitted waves in far-field region, an experimental data, the *y* component of Poynting vector (intensity detected by probes), is to be utilized and a proper expression needs to be established relating it to the various wave components. First of all, Poynting vector can be obtained starting from Maxwell's equations and the relationship of electromagnetic fields based on the interested structure. The structure is described in Fig. 37. The substrate is quartz (or SiO_2), and a thin Ag film was deposited on the substrate. After etching a single nanoslit on the thin Ag layer, where maybe a deep trench was also etched into the

substrate to make sure the complete etching of Ag film, the structure is ready for the measurement of transmitted intensity. Due to the probe scanning along the film surface located vertically, the *y*-component of Poynting vector is good enough to describe the intensity detected by the probe. The wavefronts of direct film-transmission are horizontally planar while those of slit-transmission are cylindrical as shown in Fig. 37. The coordinate relations can be written as $\cos\theta = x/r$, $\sin\theta = y/r$, where $r = \sqrt{x^2 + y^2}$, and the basis $\hat{\mathbf{r}} = \hat{\mathbf{x}}\cos\theta + \hat{\mathbf{y}}\sin\theta$, and the wave vector and its components are $k_x = k\cos\theta$, $k_y = k\sin\theta$, and $\mathbf{k}\cdot\mathbf{r} = k_x x + k_y y$. Thus, the magnetic and electric fields of direct film-transmission can be denoted as

$$\mathbf{H}_1 = \hat{\mathbf{z}} H_1 \exp(iky + i\phi_1), \tag{12}$$

$$\mathbf{E}_1 = -\hat{\mathbf{x}}\eta H_1 \exp(iky + i\phi_1). \tag{13}$$

Thus, the Poynting vector and its *y* component of direct film-transmission can be obtained directly from the definitions:

$$\langle \mathbf{S}_1 \rangle = \frac{1}{2}\mathrm{Re}[\mathbf{E}_1 \times \mathbf{H}_1^{\,*}] = \hat{\mathbf{y}}\frac{\eta}{2}H_1^2, \tag{14}$$

$$\langle S_{1y} \rangle = \frac{\eta}{2}H_1^2 = \langle S_1 \rangle. \tag{15}$$

The magnetic and electric fields of slit-transmission can be denoted as

$$\mathbf{H}_2 = \hat{\mathbf{z}}\frac{H_2}{\sqrt{r}}\exp(i\mathbf{k}\cdot\mathbf{r} + i\phi_2), \tag{16}$$

$$\begin{aligned}\mathbf{E}_2 &= -\frac{1}{i\omega\varepsilon}\nabla\times\mathbf{H}_2 = -\frac{1}{i\omega\varepsilon}\nabla\left(\frac{H_2}{\sqrt{r}}\exp(i\mathbf{k}\cdot\mathbf{r} + i\phi_2)\right)\times\hat{\mathbf{z}} \\ &= (\hat{\mathbf{r}}\times\hat{\mathbf{z}})\frac{H_2}{i\omega\varepsilon}\left(\frac{1}{2r\sqrt{r}} - \frac{ik}{\sqrt{r}}\right)\exp(i\mathbf{k}\cdot\mathbf{r} + i\phi_2) \\ &= (-\hat{\mathbf{x}}\sin\theta + \hat{\mathbf{y}}\cos\theta)\eta\frac{H_2}{\sqrt{r}}\exp(i\mathbf{k}\cdot\mathbf{r} + i\phi_2)\end{aligned}, \tag{17}$$

where the first term in the bracket in the second line in Eq. (17) is neglected at far-field because of the extra linear increase of r in the denominator. In the same way, the Poynting vector and its y-component of slit-transmission are

$$\langle\mathbf{S}_2\rangle = \frac{1}{2}\mathrm{Re}[\mathbf{E}_2\times\mathbf{H}_2^{\,*}] = (\hat{\mathbf{y}}\sin\theta + \hat{\mathbf{x}}\cos\theta)\frac{\eta}{2}\frac{H_2^2}{r} = \hat{\mathbf{r}}\frac{\eta}{2}\frac{H_2^2}{r}, \tag{18}$$

$$\langle S_{2y}\rangle = \frac{\eta}{2}\sin\theta\frac{H_2^2}{r} = \frac{\eta}{2}\frac{y}{r}\frac{H_2^2}{r}. \tag{19}$$

If we integrate $\langle\mathbf{S}_2\rangle$ passing the unit area along the half circle (as the slit-transmission wavefront, shown in Fig. 37), and set the variable of integration is dl which is tangent to the half circle (in the z direction, choosing the unit size), the power passing through the slit P_{slit} can be yielded as

$$P_{slit} = \int_{half\ circle} \langle \mathbf{S}_2 \rangle \cdot \hat{\mathbf{r}} dl = \int_0^{\pi} \frac{\eta}{2} \frac{H_2^2}{r} r d\theta = \frac{\eta\pi}{2} H_2^2 . \tag{20}$$

At far-field, the surface waves are not counted, and the Poynting vector is calculated as

$$\begin{aligned} \langle \mathbf{S} \rangle &= \frac{1}{2} \mathrm{Re}[(\mathbf{E}_1 + \mathbf{E}_2) \times (\mathbf{H}_1 + \mathbf{H}_2)^*] \\ &= \frac{\eta}{2} \left[\begin{array}{l} \hat{\mathbf{y}} H_1^2 + \hat{\mathbf{y}} \frac{H_1 H_2}{\sqrt{r}} \cos(\mathbf{k} \cdot \mathbf{r} - ky - \phi) \\ + (\hat{\mathbf{y}} \sin\theta + \hat{\mathbf{x}} \cos\theta) \frac{H_1 H_2}{\sqrt{r}} \cos(\mathbf{k} \cdot \mathbf{r} - ky - \phi) + (\hat{\mathbf{y}} \sin\theta + \hat{\mathbf{x}} \cos\theta) \frac{H_2^2}{r} \end{array} \right], \\ &= \frac{\eta}{2} \left[\hat{\mathbf{y}} H_1^2 + (\hat{\mathbf{y}} + \hat{\mathbf{r}}) \frac{H_1 H_2}{\sqrt{r}} \cos(\mathbf{k} \cdot \mathbf{r} - ky - \phi) + \hat{\mathbf{r}} \frac{H_2^2}{r} \right] \end{aligned} \tag{21}$$

where $\phi = \phi_1 - \phi_2$ is the phase retardation of slit-transmission compare to direct film-transmission. The *y* component of the Poynting vector is

$$\langle S_y \rangle = \frac{\eta}{2} \left[H_1^2 + \left(1 + \frac{y}{r} \right) \frac{H_1 H_2}{\sqrt{r}} \cos(kr - ky - \phi) + \frac{y}{r} \frac{H_2^2}{r} \right]. \tag{22}$$

With Eq. (15) and Eq. (19), the *y* component of the Poynting vector can be expressed as

$$\langle S_y \rangle = \langle S_{1y} \rangle + \left(\sqrt{\frac{r}{y}} + \sqrt{\frac{y}{r}} \right) \sqrt{\langle S_{1y} \rangle \langle S_{2y} \rangle} \cos(kr - ky - \phi) + \langle S_{2y} \rangle . \quad (23)$$

It should be noted that $<S_{2y}>$ is dependent of the location y and r from Eq. (19), which is denoted as power density at the certain location. Or with Eq. (15) and Eq. (20), the y component of the Poynting vector can be expressed as

$$\langle S_y \rangle = \langle S_1 \rangle + \left(1 + \frac{y}{r} \right) \sqrt{\frac{\langle S_1 \rangle P_{slit}}{\pi r}} \cos(kr - ky - \phi) + \frac{y}{r} \frac{P_{slit}}{\pi r}, \quad (24)$$

which is equivalent to Eq. (22). In Eq. (24), P_{slit} is independent on the location y and r, which is similar with H_2. Also, P_{slit} is the result of the integration of power density $<\mathbf{S}_2>$, so the unit is watt/meter instead of watt/meter^2 (in SI unit) which is the unit of power density $<\mathbf{S}_1>$ or $<\mathbf{S}_2>$. The unit ratio of P_{slit} over $<S_1>$ is thus meter. Similarly, the unit of H_2/H_1 is (meter)$^{1/2}$ by comparing Eq. (15) and Eq. (20).

When a probe scanning is performed, $<S_y>$ in Eq. (22) or Eq. (24) is obtained for different values of x. By choosing two points, e.g., one the central peak and another one the closest valley to the center (Fig. 35), two independent equations are obtained from Eq. (22) or Eq. (24). From the two equations, the ratio of amplitudes or powers can be solved uniquely. We

can first refer to Eq. (22) to solve the amplitude ratio because the power ratio can be easily achieved directly by the result of amplitude ratio.

For a certain y value, e.g. $y = y_0$, at $x = 0$ (central position) and x_0 (closest valley position), Eq. (22) can be recast to

$$\left\langle S_y \right\rangle (x = 0, y = y_0) = \frac{\eta}{2}\left[H_1^2 + \frac{2H_1 H_2}{\sqrt{y_0}}\cos\phi + \frac{H_2^2}{y_0} \right], \tag{25}$$

$$\left\langle S_y \right\rangle (x = x_0, y = y_0) = \frac{\eta}{2}\left[H_1^2 + \left(1 + \frac{y_0}{r_0}\right)\frac{H_1 H_2}{\sqrt{r_0}}\cos(kr_0 - ky_0 - \phi) + \frac{y_0}{r_0}\frac{H_2^2}{r_0} \right], \tag{26}$$

respectively, where $r_0 = \sqrt{x_0^2 + y_0^2}$. By comparing Eq. (25) and Eq. (26), the ratio of amplitudes between slit-transmission and direct film-transmission can be solved by

$$\frac{H_2}{H_1} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}, \tag{27}$$

where

$$A = \frac{y_0}{r_0^2} - \frac{s}{y_0}, \tag{28}$$

$$B = \left(1 + \frac{y_0}{r_0}\right)\frac{1}{\sqrt{r_0}}\cos(kr_0 - ky_0 - \phi) - \frac{2s}{\sqrt{y_0}}\cos\phi\,, \tag{29}$$

$$C = 1 - s\,, \tag{30}$$

and

$$s = \frac{\langle S_y \rangle (x = x_0, y = y_0)}{\langle S_y \rangle (x = 0, y = y_0)}. \tag{31}$$

After the verification process, the ambiguous sign in Eq. (27) has been chosen as "–" for a reasonable value.

To verify the model discussed above, simulation and experimental data are selected. Because the probe detected intensity is dependent of the radiation angle measured from the center of the slit for experimental data (Fig. 38), we have to take into account the probe effect for the selection of the intensity in different radiation angle, e.g. choosing the points of central peak and the closest valley.

By choosing a certain y value, e.g. y = 7.6 μm, the ratio of amplitudes between slit-transmission and direct film-transmission is calculated. The ratio of powers between them can also be easily obtained by applying $P_{slit}/\langle S_1 \rangle = \pi(H_2/H_1)^2$ from Eq. (15) and Eq. (20). The radiation angle of the closest valley is $90^o - \theta = \tan^{-1}(x_0/y_0) \sim 12^o$. By looking up Fig. 38 with the

radiation angle, the probe effect with intensity factors is considered into the experimental data. The results are listed in Table 3.

By reading the peak positions and values of H_z for both direct film-transmission and slit-transmission directly from FDTD simulation and setting the unit of distance r with micrometer (μm) (Fig. 39), the amplitudes of fields are estimated to be $H_1 = 0.15$, $H_2 = 0.094$ $\mu m^{1/2}$ (assuming incident $H_0 = 1$), so $H_2/H_1 = 0.6$ $\mu m^{1/2}$, and $P_{slit}/\langle S_1 \rangle = \pi(H_2/H_1)^2 = 1.13$ μm, which well matches the results extracted from experimental data.

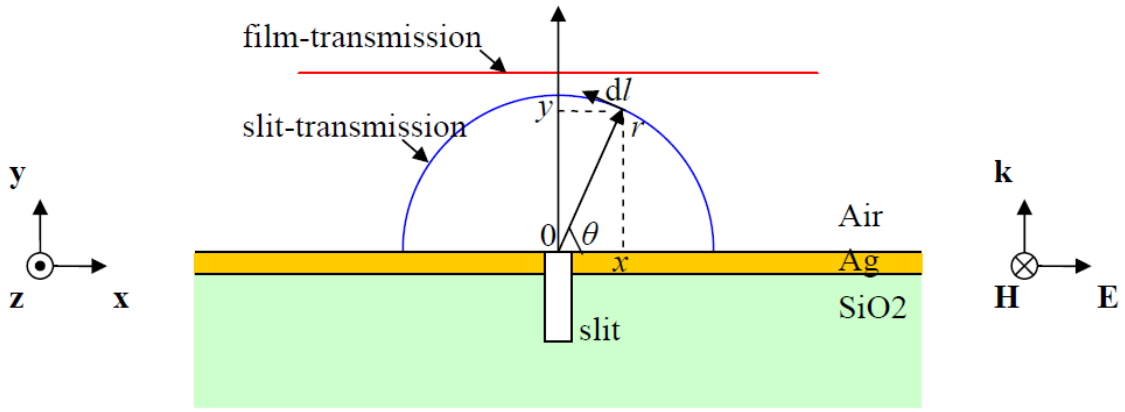


Figure 37. Wavefronts of direct film-transmission and slit-transmission. The structure is described as: the thickness of Ag is 50 nm, the width of slit of Ag and trench in quartz is 80 nm, and the depth of trench in quartz is 100 nm. The incident plane wave source is 633 nm.

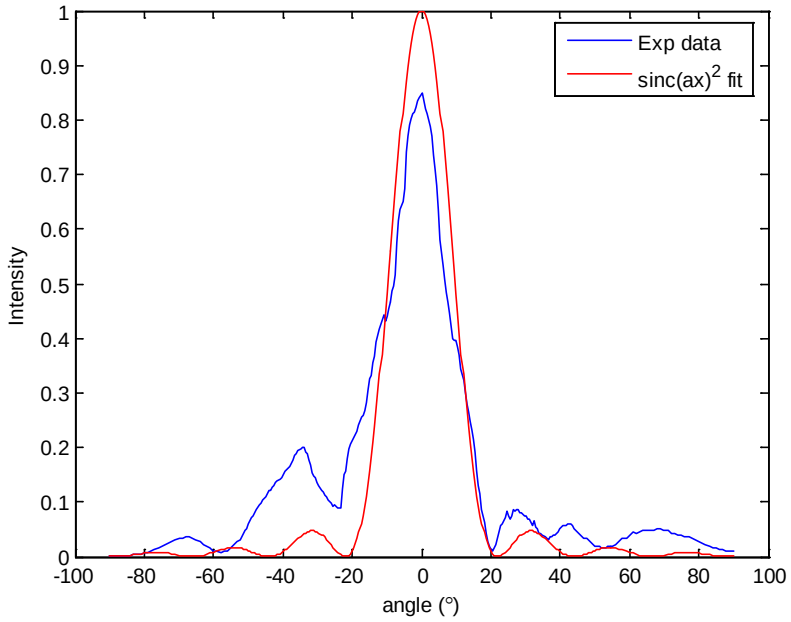


Figure 38. Intensity of probe angled detection and corresponding curve fitting.

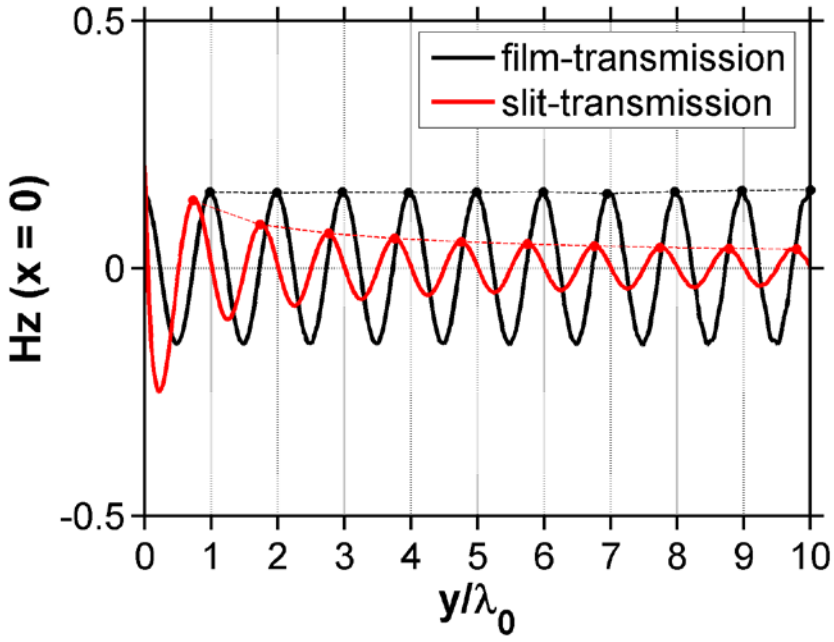


Figure 39. H_z fields at central position (x = 0) along y direction for both direct film-transmission and slit-transmission. The peaks of the fields are dotted and lined together by dashed.

Table 3. Amplitudes extraction in simulation and experiment for y = 7.6 μm after probe correction

	Simulation	**Experiment**	
		L (x_0 < 0)	**R (x_0 > 0)**
x_0 (μm)	1.64	-1.85	1.35
y_0 (μm)	7.6	7.6	7.6
$<S_y>(x = 0, y = y_0)$	0.013	0.45/0.85*	0.45/0.85*
$<S_y>(x = x0, y = y_0)$	0.0072	0.12/0.37*	0.13/0.40*
φ (º)	-88º	-86º	-86º
H_2/H_1 ($\mu m^{1/2}$)	0.65	0.62	0.60
$P_{slit}/<S_1>$ (μm)	1.33	1.19	1.13

* The denominator is the probe affected intensity factor directly read from Fig. 38.

3.3 SUMMARY

In summary, we have experimentally characterized the phase evolution of optical wavefronts emanating from a nanoslit (80-nm width) formed in a 50-nm-thick Ag film. A planar wave, directly transmitted through the thin metal film, was used as a reference in forming an interference pattern with the slit-transmitted free-space radiation and surface plasmons in the near- to far-field regimes. The phase relationship of the slit-transmitted waves with respect to the direct transmission was quantitatively established by comparing the measurement result with the analytical and FDTD simulation results.

The amplitude relationship of the slit-transmitted waves and the direct film-transmitted waves was also extracted from both experimental measurement and FDTD simulation results. The amplitudes were also directly read from the FDTD result and compared with the extracted data. Combining with the phase relationship information, complete imaging of optical wavefronts from a metal nanoaperture structure in the near- to far-fields can be achieved. This phase/amplitude imaging is expected to be important in designing advanced nano-optic and plasmonic structures where precise control of optical phase is essential.

4.0 NEGATIVE-REFRACTION DIFFRACTIVE OPTICS

Refraction of light at an interface of two different media forms an essential basis in imaging and beam-shaping optics. While commonly viewed as a macroscopic phenomenon occurring at a media interface, at a microscopic level, the phenomenon involves diffractive transmission of light through atomic or molecular level scatterers (re-radiators) and subsequent interference among the produced wavelets [5]. In the case of an interface with an artificial medium whose refractive index is negative, light can be bent to a negative angle with the surface normal. However, negative-index metamaterials commonly involve resonant structures designed at a subwavelength scale, and are intrinsically associated with loss and limited spectral width of operation [48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64]. Here we report an alternative method of negatively refracting light without use of negative-index metamaterial. An array of nanoapertures formed in a metal film is utilized in redirecting an incident beam into the negative refraction direction. Each nanoaperture is tilt-oriented from the film surface such that the radiation pattern directs to a particular orientation off from the substrate normal with reduced angular spreading. As an array, the aperture radiations make constructive interference into a specific direction that is designed to match the radiation patterns of individual apertures. This grated thin-film structure primarily supports the -1^{st} order diffraction on the transmission side (negative refraction), while the zero-order and other higher order diffractions are suppressed for a wide range of incidence angle. Unlike the bulk metamaterials case, the negative refraction achieved with the grated metal films can reach the far-field region without propagation loss and can operate at arbitrary wavelengths. This grating diffraction with order-selection capability,

enabled by engineering the radiation patterns of nanoapertures, provides access to angular ranges that are not attained in conventional optics. The nanoaperture arrays with designer radiation patterns offer a new approach to overcoming the limits of conventional refractive optics.

An intrinsic connection is evident between refraction and grating diffraction in that both phenomena involve diffractive transmission and interference. Yet, their differences lie at the vastly different length scales involved. That is, refraction occurs due to the atomic/molecular level spacing of scatterers and grating diffraction occurs due to wavelength scale aperture spacing. In the former case the wavelets produced by scatterers constructively interfere only to the zero-order direction, that is, to the direct transmission direction. In the latter case higher order diffraction beams can also be produced through interference, depending on the ratio of grating period to wavelength. In conventional gratings, however, the transmitted power is mostly carried by the zero-order diffraction, and other higher-order diffraction is usually of minor intensity. In the grating structure reported in this letter, the nanoaperture array is designed to support only the -1^{st} order diffraction on the transmission side, that is, the diffraction to the negative refraction direction, and the zero-order and other higher-order diffractions are suppressed. This negatively-refracting grating diffraction is enabled by tilting the nanoaperture orientations from the film surface such that the aperture radiation patterns become highly directed and constructively interfere only in the direction of a selected diffraction-order.

4.1 FUNDAMENTAL DESIGN: SINGLE VERTICAL NANOSLIT

The radiation pattern of a single vertical nanoslit structure was first simulated with finite-difference time-domain (FDTD) method. For a planar wave normally incident from the bottom

side, the nanoslit transmission shows a radiation pattern tilt-oriented from the substrate normal (Fig. 40(a)-(f)). For the case of a vertical slit formed on a Ag film (200-nm metal thickness; 760-nm tilt span; 260-nm step height; which corresponds to 60-nm slit width and 19^o tilt of bottom surface), for example, the main lobe is oriented to $\varphi = -50^o$ tilt direction from the substrate normal with a full-width-at-half-maximum (FWHM) angle of 50^o. This is in a striking contrast with the radiation pattern of a conventional slit whose dipole axis is horizontal, parallel to the film surface (Fig. 40(e), (f)). The latter structure shows a nearly uniform distribution of power for radiation angle φ of -70^o to $+70^o$. This tilt-nanoslit aperture configuration serves as a dipole-like source that oscillates perpendicular to the metal film under excitation by an incident wave. For a transverse magnetic (TM)-polarized light incident to an aperture, polarization charges of opposite polarities are induced around the opposing corners (edges) of nanoslit (Fig. 40(g)-(i)). The dipole oscillation at slit edges has the effect of re-radiating incident energy primarily into the direction perpendicular to the dipole axis, that is, towards the horizontal direction in the vertical nanoslit case. The Poynting vector distribution (Fig. 40(i)) clearly reveals the overall energy flow through the vertical nanoslit structure. The incident wave is reflected back at the bottom faces of both slabs, forming a standing wave. In the corner region a vortex of turbulent energy flow develops. The boundary diffraction at corner edges enables funnelling of the incident power that falls in the near-field region of the corners [36]. As a result, part of the energy flow around the vortex leaks through the gap and radiates away on the exit side. While most of the slit-transmitted power radiates into free space, a certain portion propagates on metal surfaces, preferentially on the horizontal surface. This surface-bound energy flow is carried by surface plasmons emanating from slit edges [1,2,26].

The amount of dipole charges induced on slit corners and edges depends on the aperture geometry/dimension and the orientation of incident field vectors with respect to the edges. An electromagnetic wave interacting with metal surface (slit edges) can induce polarization surface charges, whose surface density can be expressed as [6] $\sigma_{\text{pol}} = \varepsilon_0(\mathbf{E}_2 - \mathbf{E}_1)\cdot\mathbf{n}_{21} = \varepsilon_0\left(1-\frac{\varepsilon_{\text{D}}}{\varepsilon_{\text{M}}}\right)\mathbf{E}_2\cdot\mathbf{n}_{21}$. Here $\mathbf{E}_1$ and $\mathbf{E}_2$ are the electric field on the metal (ε_{M}) and dielectric (ε_{D}) side of the interface, respectively, and $\mathbf{n}_{21}$ is the normal vector to the surface. ε_0 is the dielectric permittivity of free space. For most metals/dielectrics, $|\varepsilon_{\text{M}}| >> |\varepsilon_{\text{D}}|$, therefore, the amount of dipole charges induced on the surface would not be critically dependent on the dielectric constant of metal. The strong vertical electric field, E_{y} present around the corner gap region (Fig. 40(h)) is caused primarily by boundary diffraction of an incident TM-wave at the bottom face of the upper slab [5,36]. This vertical field component around the corner induces surface charges on the horizontal faces in the gap region. Charge induction in the corner region involves a different process. Reflection of an incident wave induces surface current $\mathbf{J}_{\text{S}} = \mathbf{n}\times\mathbf{H}$, where **H** is the magnetic field on the surface. This surface current then induces local charges at the corner governed by the continuity equation. The strong local charges induced at the bottom corner of the upper slab, for example, are due to the surface current induced by the incident TM fields (Fig. 40(i)). The resulting surface charge density at the corner can be expressed as $\sigma_{\text{pol}} = \frac{1}{i\omega}J_S = \frac{1}{i\omega}(1-r)H_{\text{in}}$. Here H_{in} is the incident magnetic field at the metal surface and r is the reflection coefficient. The reflection coefficient remains close to -1 for most metals, and this implies negligible dependence of the dipole charge density on metal's dielectric constant. Overall this analysis suggests that vertical nanoslits made of metals with different dielectric constants would show radiation patterns of similar strength (*i.e.*, transmission throughput).

Fig. 41(a) shows a measurement result of optical transmission (at 633 nm wavelength) through a vertical nanoslit (~100-nm slit width) formed in a 100-nm-thick Cr film. The sample was prepared by angle deposition of a Cr film on a quartz substrate, whose surface was focused-ion-beam (FIB) etched to an asymmetric saw-tooth profile with one side vertically step-etched (step height of ~200 nm) and the other side taper-etched over a tilt span of 760 nm (see Methods for details). The radiation pattern was measured by scanning a near-field scanning optical microscope (NSOM) probe in the near to far field region. The sensitivity of a nanoapertured NSOM probe varies depending on beam incidence angle, that is, showing lower sensitivity for larger incidence angle [22,24]. In this work, scanning was performed with the probe axis tilt-oriented to the peak radiation direction (Fig. 41(a)) for maximum signal strength. The NSOM-scan result reveals -45^{o} tilt of the main lobe, well matching the simulation result (Fig. 41(b)).

The transmission throughput of a vertical nanoslit was also calculated for different incidence angles (Fig. 42). While the radiation pattern remains tilt-oriented at the same angle ($\varphi = -50^{o}$) for varying incidence angles, the throughput transmission changes significantly. The transmission efficiency, defined as the total transmitted power (both free-space radiation and surface plasmon power) divided by the incident power that falls on the slit cross-section, reaches 80 % level at $\theta = +45^{o}$ incidence or normal incidence angle, and 350 % at $\theta = -45^{o}$ incidence. Here the transmission efficiency's being greater than 100 % indicates a funnelling effect, that is, the nanoslit efficiently captures incident power and transmits more than the amount of power that falls on the aperture area. For the case of a horizontal slit with same metal thickness and slit width, the transmission efficiency reaches maximum 53 % at normal incidence. This implies that a vertical nanoslit structure can outperform a horizontal nanoslit in transmitting incident power. For thicker films, this performance contrast becomes even stronger, and this trend can be

understood in view of the fact that in horizontal nanoslits a significant degree of transmission loss occurs in the narrow channel region whereas in the vertical slit case there is no such waveguide constriction and therefore no attenuation [32,65]. The high-throughput transmission and highly-directed radiation pattern of a vertical nanoslit structure offers interesting potential for overcoming the limits of conventional refractive optics.

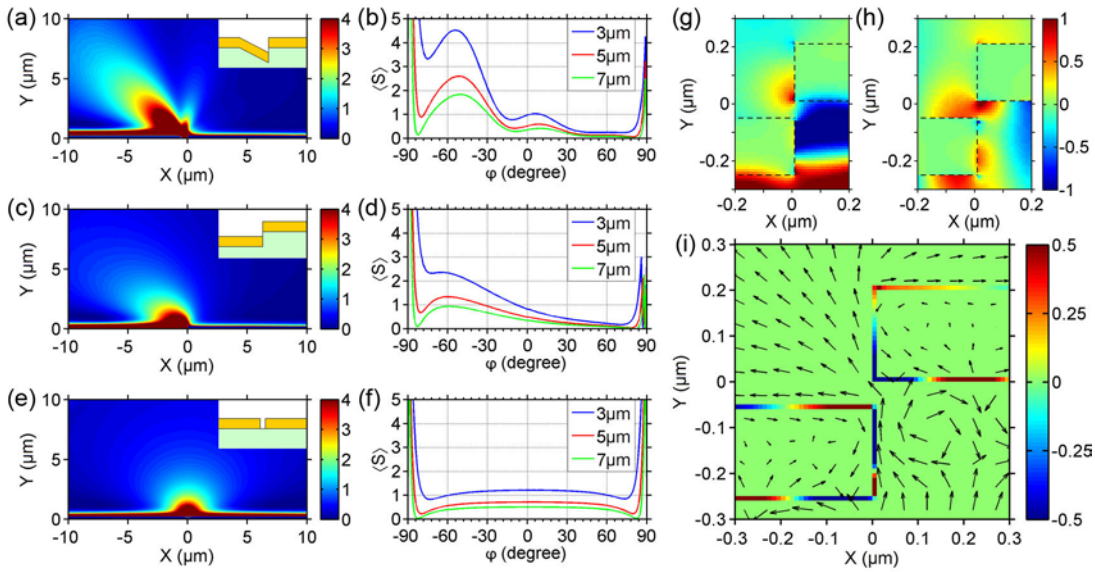


Figure 40. Radiation patterns of vertical nanoslit apertures. FDTD simulation of optical transmission through single-nanoslit structures formed on a Ag layer (60-nm slit width, 200-nm metal thickness). The nanoslit dipole axis is either vertically (a & b; c & d) or horizontally (e & f) oriented to the substrate (see the insets for schematic cross-sections). The width of the slanted portion of metal in a is 760 nm. A TM-polarized light (633 nm wavelength) is normally incident from the bottom side. Poynting vector distributions on the exit side: magnitude maps (a, c, and e) and angular profiles of magnitude scanned at three different radial distances 3 μm, 5 μm, or 7 μm (b, d, and f). The radiation angle φ refers to the substrate normal. Electric field distributions: Snapshot images of E_x (g) and E_y (h) around the corner gap region. Time-averaged energy flow distribution: A log scale plot of Poynting vectors (i). Note the presence of arrows of significantly smaller length inside the metal and they orient to the opposite directions to the ones in the air side along the interface. A surface polarization charge distribution that corresponds to the field distributions (g, h) is also shown on the metal/dielectric interfaces (i).

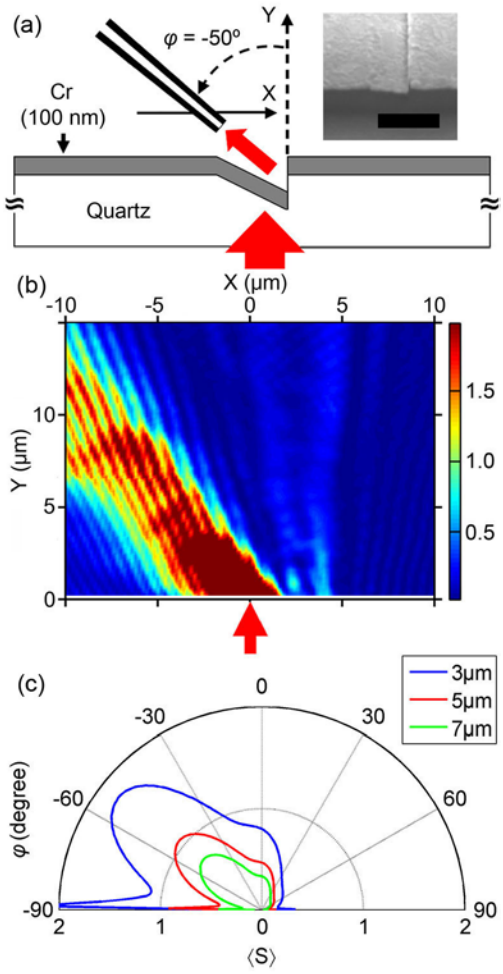


Figure 41. Measured radiation pattern of a vertical nanoslit. (a) Scanning probe measurement of optical transmission through a metal nanoslit (inset: SEM image of a vertical nanoslit sample; scale bar, 1 μm). The probe axis was tilt-oriented to the peak radiation direction ($\varphi = -50^{o}$). (b) A measured beam profile of a Cr vertical nanoslit (100-nm slit width; 100-nm Cr thickness) for normal incidence of TM-polarized light (633 nm wavelength). (c) FDTD simulation of radiation patterns of the Cr nanoslit sample calculated at three different radial distances from the aperture center.

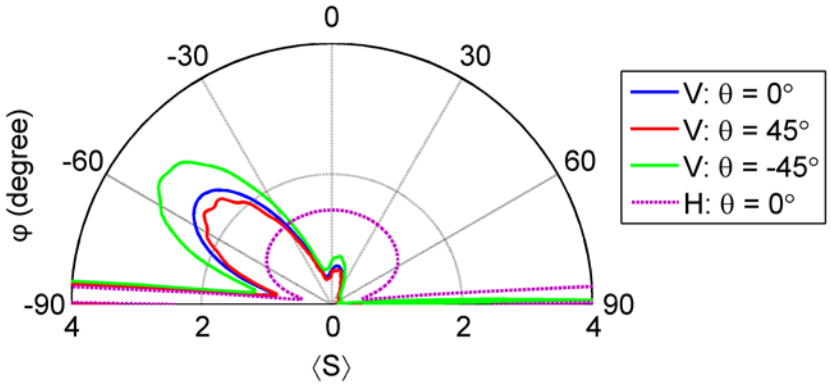


Figure 42. Radiation patterns of a vertical nanoslit calculated for different incidence angles. Sample structure: 200-nm metal (Ag) thickness; 760-nm slant span; 60-nm slit width. A TM-polarized light (633 nm) is incident to the substrate at three different angles θ = +45^o (red), 0^o (blue: the intensity scaled down by a factor of 2) or -45^o (green: the intensity scaled down by a factor of 4). For comparison, a horizontal nanoslit (200-nm-thick Ag; 60-nm slit width) is also shown (dotted, purple). All radiation patterns were calculated at r = 5 μm radial distance.

4.2 NEGATIVE REFRACTION ENABLED BY THE VERTICAL DIPOLE ARRAY

An array of vertical dipole nanoslit was developed in the form of a periodically-slanted metal film and was investigated for the possibility of redirecting an incident energy flow into the negative refraction direction (Fig. 43(a), (c)). The grating period is determined such that the aperture radiations make constructive interference into a negative refraction direction for a given wavelength of light. Being an interference phenomenon, the negative refraction at a vertical-dipole nanoaperture array is also governed by the Bragg law of diffraction gratings: $\sin\varphi = \sin\theta + m\lambda/d$. Here θ is the incidence angle measured on the air side of substrate, and φ is the refraction angle measured on the exit side (air). d is the grating period, and λ is the free space wavelength. m is an integer representing the diffraction order. For the case of negative refraction being discussed in this work, m = -1. The refraction relationship is then expressed as $\sin\varphi = \sin\theta - \lambda/d$. This formula can be viewed as *the Snell's law for negative refraction* in the sense that it relates incidence angle to refraction angle at an interface of two positive-index media. Note that in this formula, no substrate effect is explicitly involved. This is because the incidence angle is defined on the air side, not inside the substrate.

Fig. 43(a), (b) shows scanning-probe measurement results of optical transmission (TM-polarized, 633 nm wavelength) through Ag nanoslit arrays (20 slits with 760-nm grating period). For comparison, an array of conventional, horizontal nanoslits with the same grating period was also fabricated and tested (Fig. 43(b)). The vertical nanoslit array clearly demonstrates negative refraction with refraction angle φ of -20^o for incidence angle θ of 30^o. This negative refraction beam corresponds to the -1st order diffraction from the grating. In contrast, the conventional nanoslit array with in-plane horizontal dipole apertures shows the zero-order transmission (direct

transmission) as a dominant beam as expected. In both cases, the scanning was performed with the probe axis aligned normal to the substrate. Considering the near symmetric angles (φ of $+30^o$ and -20^o) of the zero-order and -1^{st} order beams with respect to the substrate normal, the probe effect is believed to be negligible in these scanning measurements. Overall the result unambiguously confirms that the vertical nanoslit array structure blocks direct transmission and supports only the -1^{st} order transmission.

This negative-refraction beam would become even stronger in its intensity/throughput when the -1^{st} order diffraction angle is designed to fall safely within the angular ranges of nanoaperture radiation patterns. In the case of radiation pattern simulated in Fig. 40, for example, the radiation angle φ ranges from -25^o to -75^o in terms of half-maximum intensity points. For the given grating period (760 nm) and wavelength (633 nm), the incidence angle needs to stay close to the range of 0^o to 25^o in order to maintain high intensity of negative refraction. Fig. 43(c) shows a measurement result with a Cr vertical nanoslit array sample (10 slits with 760-nm period) tested at different incidence angles. The probe axis was aligned to the peak radiation direction estimated from the Bragg condition. The scanned images reveal well-defined, straight beams that extend far beyond 15-20 μm distance from the array. At normal incidence, refraction angle of -55^o is observed (Fig. 43(c), middle). At $\theta = +15^o$ incidence, $\varphi = -35^o$ was observed (bottom). Again both results agree well with the Bragg condition. For the case of smaller grating period, e.g., 500 nm, the incidence angle can be in a relatively wider range, that is, 20^o to 60^o for half-maximum transmission points. The corresponding refraction angle would then be in the range of -75^o to -25^o. A measurement result with a 500-nm-period Cr vertical nanoslit array sample also demonstrates negative refraction governed by the Bragg law (Fig. 44). In general, Ag samples showed stronger transmission compared to Cr samples. This enhanced transmission

of Ag samples is attributed to the surface plasmons that are excited at slit edges, propagating to neighbor slits and there contributing to dipole radiation to free space [12,13,66].

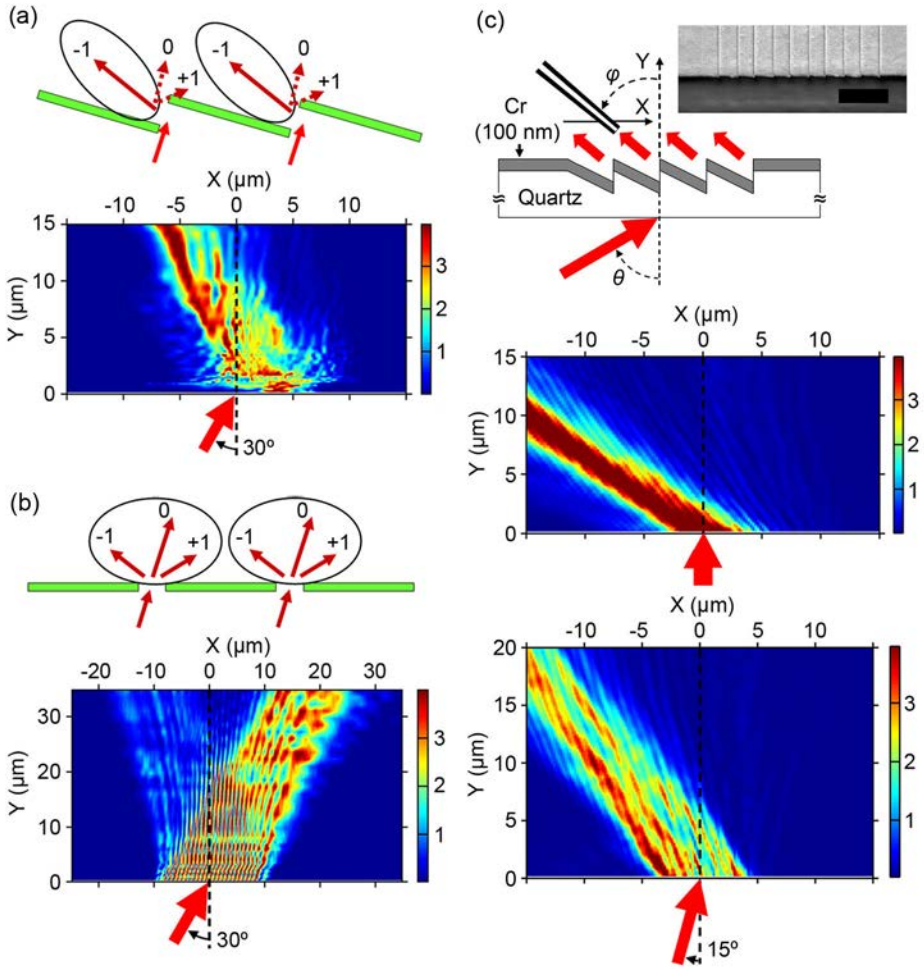


Figure 43. Negative refraction of light at vertical nanoslit arrays. (a) A measured beam profile of a Ag, vertical-nanoslit array (20 slits with 760 nm period). Negative refraction is observed with refraction angle φ = -20^{o} for incidence angle θ = 30^{o}, which corresponds to the -1st order diffraction from the grating. (b) A measured beam profile of a conventional, horizontal-nanoslit array (20 slits with 760 nm period) formed on a Ag layer. The zero-order, direct transmission is observed as a primary beam for the same incidence angle as above. In both cases (b, c), the probe was aligned normal to the substrate. (c) Negative refraction with a Cr, vertical-nanoslit array (10 slits with 760 nm period): measurement setup (top; inset, SEM image of a Cr sample; scale bar, 2 μm); Measured beam profiles for normal incidence (middle) or for incidence angle θ = 30^{o} (bottom). Refraction angle φ = -55^{o} or -35^{o} was observed, respectively.

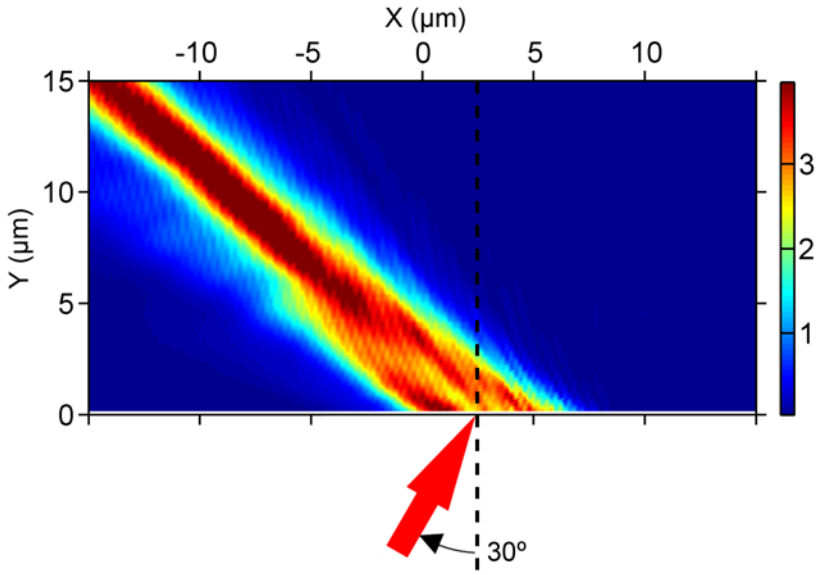


Figure 44. Measured beam profile of a Cr, vertical-nanoslit array (15 slits with 500 nm period). A TM-polarized light (633 nm) is incident to the substrate at $\theta = +30^{o}$ angle.

4.3 SUMMARY

This negative refraction function offers a new mode of imaging of rear views in a side-forward direction. We note that this negative-refraction mirror imaging does not incur the conventional ‘mirror imaging’ effect, that is, there would be no inversion of the right-left relationship of an image. In conventional mirror imaging, the ray components of the right and left sides of an image travel the same distance. Therefore, as the beam propagation direction is reversed upon mirror reflection, the right-left relationship of a mirror image becomes flipped over compared to that of the original image seen directly without a mirror. In the case of negative-refraction mirror imaging, the ray components in the outer side of a beam travel progressively longer distances compared to the inner-side ray components. This optical path length difference for the ray components on the transverse domain provides the phase retardations that are required to offset the image inversion effect incurred with mirror reflection, that is, with reversal of beam propagation direction. This negative refraction with non-inverting mirror-imaging function, enabled by grating diffraction with order-selection capability, offers interesting potential for beam-shaping and imaging optics.

5.0 CONCLUSION

In this thesis, we have investigated metallic wedges and nanostructures as a medium for plasmon interactions. Surface plasmon waves excited by a metallic wedge or nanostructure can produce many interesting phenomena that can be simulated and observed in the near-field to far-field regime. The scope of this study covers both theoretical (analytical and simulation) and experimental aspects. The metallic wedges studied in this thesis include a single wedge with or without thin film. The incident light is along one surface of the wedge with magnetic polarization. The metallic nanostructures studied in this thesis include nanoslit array with variable number of slits, ranging from single, to 2-10 or even more. The spatial extent of SP fields is usually an order of magnitude smaller than the optical wavelength. Therefore, the structures are designed to span nano to micrometer length scales in order to accommodate wave interactions at different length scales, i.e., the slit width of the 10 nm order and the grating period of the 100 nm order.

We report a theoretical and experimental study of electromagnetic interactions of a two-dimensional (2D) metal wedge with two semi-infinite surfaces with a planar wave normal incident to one of the two surfaces. By employing the FDTD method we demonstrate the essence of a boundary diffraction wave which is produced due to the discontinuity of structure and in conjunction with reflection from the front and vertical surface. We analytically show the metal wedge structure diffracts the incident wave by setting up a model based on a series of line sources. We show that a metal corner structure possesses an intrinsic capability of directing the

incident power around the edge into the forward direction. The interplay of the boundary diffraction wave and the incident and reflection waves in the near field region of a metal corner is found to form a basis of the funneling phenomena that are commonly observed in metal nanoslit structures. The incident wave propagating parallel to the sidewall is also found to experience destructive interference with the boundary diffraction waves and thus to form a depleted-energy-flow region along the glancing angle direction. With the model, we derive the boundary diffraction waves and the total power on the front surface of the wedge at the air side. With the simulation and measurement results of the corresponding power, we extract the phase change of reflection by the metal for the given incident wave. From the optical measurement of the phase change, we can obtain the refractive index of metals.

We discuss near- to far-field imaging of phase evolution of optical wavefronts emanating from a nanoslit formed in a thin (50 nm thick) Ag film. The evolution of optical phases is imaged using a self-interference technique in conjunction with a scanning probe method. The phase relationship of the slit-transmitted waves with respect to the direct transmission through the thin metal film is quantitatively established. The singular-phase points resulting from the interplay of slit diffraction and surface plasmons are identified in the intermediate-field region. Imaging of optical wavefronts in the near- to far-fields is expected to be important in designing advanced nano-optic and plasmonic structures where precise control of optical phase is essential.

Finally, we report negative refraction of visible light enabled by a nanostructured thin film grating that is placed in conventional positive-index media, that is, at a silica/air interface. The negative refraction via grating diffraction can operate at arbitrary wavelengths. By locally varying the grating period and/or the tilt angle of individual aperture, the transmission profile can also be adjusted to match the incident beam profile. The interfacial negative refraction without

bulk media loss offers a promising approach to accessing angular ranges that have not been reachable in conventional optics.

BIBLIOGRAPHY

[1] Raether, H. *Surface Plasmons on Smooth and Rough Surfaces and on Gratings* (Springer-Verlag, Berlin, 1988).

[2] Barnes, W. L., Dereux, A. & Ebbesen, T. W. Surface plasmon subwavelength optics. *Nature* **424**, 824-830 (2003).

[3] Taflove, A. & Hagness, S. C. *Computational Electrodynamics: the Finite-difference Time-domain Method*. (Artech House, 2005).

[4] Chaillou, S., Wiart, J. & Tabbara, W. A subgridding scheme based on mesh nesting for the FDTD method. *Microwave and Opt. Techn. Lett.*, **22**, 211-214 (1999).

[5] Born, M. & Wolf, E. *Principles of Optics*, 7th ed. (Cambridge University Press, Cambridge, 1999).

[6] Jackson, J. D. *Classical Electrodynamics*, 3rd ed. (Wiley, Hoboken, 1999).

[7] Sommerfeld, A. *Optics* (Academic Press, New York, 1954).

[8] Rubinowicz, A. Thomas Young and the theory of diffraction. *Nature* **180**, 160-162 (1957).

[9] Keller, J. B. Geometrical Theory of Diffraction. *J. Opt. Soc. Am.* **52**, 116-130 (1962).

[10] Umul, Y. Z. Alternative interpretation of the edge-diffraction phenomenon. *J. Opt. Soc. Am. A* **25**, 582-587 (2008).

[11] Lezec, H. J. *et al.*, Beaming light from a subwavelength aperture. *Science* **297**, 820-822 (2002).

[12] Lalanne, P. & Hugonin, J. P. Interaction between optical nano-objects at metallo-dielectric interfaces. *Nature Phys.* **2**, 551-556 (2006).

[13] López-Tejeira, F. *et al.*, Efficient unidirectional nanoslit couplers for surface plasmons. *Nature Phys.* **3**, 324-328 (2007).

[14] Betzig, E., Harootunian, A., Lewis, A. & Isaacson, M. Near field diffraction by a slit: implications for superresolution microscopy. *Appl. Opt.* **25**, 1890-1900 (1986).

[15] Kukhlevsky, S. V., Mechler, M., Csapó, L., Janssens, K. & Samek, O. Enhanced transmission versus localization of a light pulse by a subwavelength metal slit. *Phys. Rev. B* **70**, 195428 (2004).

[16] Seo, M. A. *et al.*, Terahertz field enhancement by a metallic nano slit operating beyond the skin-depth limit. *Nature Photon.* **3**, 152-156 (2009).

[17] Palik, E. D. ed., *Optical Constants of Solids*. (Academic Press, New York, 1998).

[18] Ufimtsev, P. Ya. *Fundamentals of the physical theory of diffraction*. (Wiley-IEEE, New Jersey, 2007).

[19] Umul, Y. Z. Modified theory of physical optics approach to wedge diffraction problems. *Opt. Express* **13**, 216-224 (2005).

[20] Hecht, E. *Optics*, 4th ed. (Addison-Wesley, San Francisco, 2002).

[21] Chen, L., Robinson, J. T. & Lipson, M. Role of radiation and surface plasmon polaritons in the optical interactions between a nano-slit and a nano-groove on a metal surface. *Opt. Express* **14**, 12629-12636 (2006).

[22] Jung, Y. S., Wuenschell, J., Schmidt, T. & Kim, H. K. Near- to far-field imaging of free-space and surface-bound waves emanating from a metal nanoslit. *Appl. Phys. Lett.* **92**, 023104 (2008).

[23] Kihm, H. W. *et al.*, Separation of surface plasmon polariton from nonconfined cylindrical wave launched from single slits. *Appl. Phys. Lett.* **94**, 141102 (2009).

[24] Jung, Y. S., Xi, Y., Wuenschell, J. & Kim, H. K. *et al.*, Near- to far-field imaging of phase evolution of light emanating from a metal nanoslit. *Opt. Express* **16**, 18875-18882 (2008).

[25] Nkoma, J., Loudon, R. & Tilly, D. R. Elementary properties of surface polaritons. *J. Phys. C: Solid State Phys.* **7**, 3547-3559 (1974).

[26] Xie, Y., Zakharian, A. R., Moloney, J. V. & Mansuripur, M. Transmission of light through slit apertures in metallic films. *Opt. Express* **12**, 6106-6121 (2004).

[27] Wuenschell, J. & Kim, H. K. Surface plasmon dynamics in an isolated metallic nanoslit. *Opt. Express* **14**, 10000-10013 (2006).

[28] Wuenschell, J. & Kim, H. K. Excitation and propagation of surface plasmons in a metallic nanoslit structure. *IEEE Trans. Nanotechnol.* **7**, 229-236 (2008).

[29] García de Abajo, F. J. Colloquium: light scattering by particle and hole arrays. *Rev. Mod. Phys.* **79**, 1267-1290 (2007).

[30] Kang, J. H., Kim, D. S. & Park, Q. H. Local Capacitor Model for Plasmonic Electric Field Enhancement. *Phys. Rev. Lett.* **102**, 093906 (2009).

[31] Takakura, Y. Optical resonance in a narrow slit in a thick metallic screen. *Phys. Rev. Lett.* **86**, 5601-5603 (2001).

[32] Bravo-Abad, J., Martin-Moreno L. & Garcia-Vidal, F. J. Transmission properties of a single metallic slit: From the subwavelength regime to the geometrical-optics limit. *Phys. Rev. E* **69**, 026601 (2004).

[33] Kihm, H. W., Lee, K. G., Kim, D. S., Kang, J. H. & Park, Q. H. Control of surface plasmon generation efficiency by slit-width tuning. *Appl. Phys. Lett.* **92**, 051115 (2008).

[34] Garcia-Vidal, F. J., Lezec, H. J., Ebbesen, T. W. & Martin-Moreno, L. Multiple Paths to Enhance Optical Transmission through a Single Subwavelength Slit, *Phys. Rev. Lett.* **90**, 213901 (2003).

[35] Aigouy, L. *et al.*, Near-Field Analysis of Surface Waves Launched at Nanoslit Apertures, *Phys. Rev. Lett.* **98**, 153902 (2007).

[36] Xi, Y., Jung, Y. S. & Kim, H. K. Interaction of light with a metal wedge: the role of diffraction in shaping energy flow. *Opt. Express* **18**, 2588-2600 (2010).

[37] Ganci, S. An experiment on the physical reality of edge-diffracted waves. *Am. J. Phys.* **57**, 370-373 (1989).

[38] Genet, C. & Ebbesen, T. W. Light in tiny holes. *Nature* **445**, 39-46 (2007).

[39] Yin, L. *et al.*, Surface plasmons at single nanoholes in Au films. *Appl. Phys. Lett.* **85**, 467-469 (2004).

[40] Gay, G. *et al.*, The optical response of nanostructured surfaces and the composite diffracted evanescent wavemodel. *Nature Phys.* **2**, 262-267 (2006).

[41] Lee, K. G. *et al.*, Vector field microscopic imaging of light. *Nature Photon.* **1**, 53-56 (2007).

[42] Ebbesen, T. W., Lezec, H. J., Ghaemi, H. F., Thio, T. & Wolff, P. A. Extraordinary optical transmission through sub-wavelength hole arrays. *Nature* **391**, 667-669 (1998).

[43] Sun, Z., Jung, Y. S. & Kim, H. K. Role of surface plasmons in the optical interaction in metallic gratings with narrow slits. *Appl. Phys. Lett.* **83**, 3021-3023 (2003).

[44] Walford, J. N., Nugent, K. A., Roberts, A. & Scholten, R. E. Three-dimensional phase imaging with a scanning optical-fiber interferometer. *Appl. Opt.* **38**, 3508-3515 (1999).

[45] Balistreri, M. L., Korterik, J. P., Kuipers, L. & Van Hulst, N. F. Local observations of phase singularities in optical fields in waveguide structures. *Phys. Rev. Lett*. **85**, 294-297 (2000).

[46] Nesci, A., Dändliker, R. & Herzig, H. P. Quantitative amplitude and phase measurement by use of a heterodyne scanning near-field optical microscope. *Opt. Lett.* **26**, 208-210 (2001).

[47] Sullivan, D. M. *Electromagnetic Simulation Using the FDTD Method*, (IEEE Press, New York, 2000).

[48] Veselago, V. G. The electrodynamics of substances with simultaneously negative values of ε and μ. *Sov. Phys. Usp.* **10**, 509-514 (1968).

[49] Pendry, J. B. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85**, 3966-3969 (2000).

[50] Notomi, M. Theory of light propagation in strongly modulated photonic crystals: Refraction-like behaviour in the vicinity of the photonic band gap. *Phys. Rev. B* **62**, R10696-R10705 (2000).

[51] Luo, C., Johnson, S. G. & Joannopoulos, J. D. Subwavelength imaging in photonic crystals. *Phys. Rev. B* **68**, 045115 (2003).

[52] Pendry, J. B. & Smith, D. R. Reversing light: negative refraction. *Phys. Today* **57**, 37-43 (2004).

[53] Zhang, S. *et al.* Experimental demonstration of near-infrared negative-index metamaterials. *Phys. Rev. Lett.* **95**, 137404 (2005).

[54] Fang, N., Lee, H., Sun, C. & Zhang, X. Sub–diffraction-limited optical imaging with a silver superlens. *Science* **308**, 534-537 (2005).

[55] Shalaev, V. M. *et al.* Negative index of refraction in optical metamaterials. *Opt. Lett.* **30**, 3356-3358 (2005).

[56] Dolling, G., Enkrich, C., Wegener, M., Soukoulis, C. M. & Linden, S. Simultaneous negative phase and group velocity of light in a metamaterial. *Science* **312**, 892-894 (2006).

[57] Hoffman, A. J. *et al.* Negative refraction in semiconductor metamaterials. *Nature Mater.* **6**, 946-950 (2007).

[58] Shalaev, V. M. Optical negative-index metamaterials. *Nature Photon.* **1**, 41-48 (2007).

[59] Lezec, H. J., Dionne, J. A. & Atwater, H. A. Negative refraction at visible frequencies. *Science* **316**, 430-432 (2007).

[60] Stockman, M. I. Criterion for negative refraction with low optical losses from a fundamental principle of causality. *Phys. Rev. Lett.* **98**, 177404 (2007).

[61] Menon, L., Lu, W. T., Friedman, A. L., Bennett, S. P., Heiman, D. & Sridhar, S. Negative index metamaterials based on metal-dielectric nanocomposites for imaging applications. *Appl. Phys. Lett.* **93**, 123117 (2008).

[62] Yao, J. *et al.* Optical negative refraction in bulk metamaterials of nanowires. *Science* **321**, 930 (2008).

[63] Shin, J., Shen, J. T. & Fan, S. Three-dimensional metamaterials with an ultrahigh effective refractive index over a broad bandwidth. *Phys. Rev. Lett.* **102**, 093903 (2009).

[64] Shumin, X. *et al.* Loss-free and active optical negative-index metamaterials. *Nature* **466**, 735-738 (2010).

[65] Sun, Z. & Kim, H. K. Refractive transmission of light and beam shaping with metallic nano-optic lenses. *Appl. Phys. Lett.* **85**, 642-644 (2004).

[66] Schouten, H. F. *et al.* Plasmon-assisted two-slit transmission: Young's experiment revisited. *Phys. Rev. Lett.* **94**, 053901 (2005).